

The acquisition of auxiliary syntax: BE and HAVE

ANNA L. THEAKSTON, ELENA V. M. LIEVEN, JULIAN M. PINE, and
CAROLINE F. ROWLAND*

Abstract

This study examined patterns of auxiliary provision and omission for the auxiliaries BE and HAVE in a longitudinal data set from 11 children between the ages of two and three years. Four possible explanations for auxiliary omission—a lack of lexical knowledge, performance limitations in production, the Optional Infinitive hypothesis, and patterns of auxiliary use in the input—were examined. The data suggest that although none of these accounts provides a full explanation for the pattern of auxiliary use and nonuse observed in children’s early speech, integrating input-based and lexical learning-based accounts of early language acquisition within a constructivist approach appears to provide a possible framework in which to understand the patterns of auxiliary use found in the children’s speech. The implications of these findings for models of children’s early language acquisition are discussed.

Keywords: language acquisition; auxiliary syntax; constructivist approach.

1. Introduction

1.1. Theoretical background

Generally speaking, English-speaking children begin to produce auxiliaries from around the age of 2;0, although adultlike competence is not achieved until much later in development, often well into the fourth year.

Auxiliaries provide one of the primary means of encoding tense and agreement and are therefore central to most theories of adult language use. However, the adult auxiliary system, and thus the question of what must be acquired, is viewed very differently by the proponents of

two opposing theoretical frameworks, the principle-based generativist approach and the usage-based constructivist approach.

Within the generativist framework, grammar is assumed to consist of a series of rule-governed processes that operate over categories of linguistic items. When constructing an utterance, speakers are assumed to retrieve individual lexical items in accordance with the syntactic rules governing the combination of words in sequence. In a generativist framework, the auxiliary system is assumed to be purely structural, devoid of semantic or functional content and therefore unlearnable in principle (Chomsky 1957). To explain language acquisition, it is necessary to credit children with a pre-given knowledge of grammar, and thus their task in acquisition is to map the words of their target language onto pre-existing abstract categories (Pinker 1989, although see Bowerman 1990). As the properties of tense and agreement are assumed to be innately specified, children's early auxiliary use is considered by many researchers to reflect underlying, adultlike grammatical competence (Valian 1991; Wexler 1994; Rice et al. 1998).

In a usage-based framework, rather than assuming that adults generate utterances by the rule-governed combination of appropriate grammatical categories, the focus is on constructions as the basic units of language (Langacker 1987; Goldberg 1995; Bybee 1998; Croft 2001). Adults are assumed to store individual utterances or concrete, lexically based constructions for use on a specific occasion *as well as* more abstract schemas, for example a general pattern for subject-auxiliary inversion. Thus, adult linguistic knowledge consists of a structured network of constructions represented at different levels of abstractness, although the nature of the syntactic and semantic relations that hold between constructions and how they are represented is viewed rather differently within different construction grammar frameworks (e.g., Lakoff 1987; Langacker 1987; Goldberg 1995; Kay and Fillmore 1999; Croft 2001). With respect to auxiliaries, Langacker (1991: 228) suggests that adults operate with a "substantial number of constructional schemas [that] constitute a highly structured assembly, in which the successive combination of simpler units gives rise to schemas of progressively greater complexity".

One central difference between the generativist and usage-based approaches to acquisition lies in the knowledge credited to the child. Generativist theories credit the child with a knowledge of grammatical categories, whereas usage-based approaches assert that the child comes to the task of acquisition with only general cognitive learning principles such as pattern extraction and categorization. Thus, grammatical categories such as Noun, Verb, or Auxiliary are thought to exist only to the extent that they can be derived from the distributional regularities of the

input (Croft 2001; Langacker 1991). The child's task in acquisition is to acquire lexically based constructions, and to build up the more abstract constructions that underlie adult language use over the course of development. From this perspective, understanding the processes behind language acquisition is central to understanding the nature of the adult end-state.

Given the importance of form–function mappings in the usage-based approach to adult language use, it is unsurprising that the child's socio-cognitive development is seen as central to the acquisition of language. Language acquisition is viewed as a drive to communicate in a way that is conventional, and thus requires the acquisition of conventional units of speech (Tomasello 2000). Children are thought to begin with lexically specific constructions that are acquired in the context of social routines, providing them with the form-function mapping central to the acquisition process (Tomasello 1992; Pine et al. 1998). The development of more abstract constructions is dependent on children recognizing relations between the form and meaning of a number of independently acquired lexically based constructions. Consequently, children's early use of auxiliaries may represent the use of lexically based constructions prior to the development of more abstract constructions.

1.2. *Auxiliary use and nonuse in early speech*

During the early stages of language acquisition, children pass through a stage in development when they produce both finite sentences containing an auxiliary (*she is sleeping*) and nonfinite sentences where the auxiliary is omitted (*she sleeping*). There is little agreement among researchers as to the reasons for this observed pattern of auxiliary use and nonuse. In the current article, first, we document the pattern of use and nonuse, in obligatory contexts, of the auxiliaries BE and HAVE in the speech of English-speaking children. BE and HAVE are the only two auxiliaries for which omission can be reliably identified and the form of the omitted auxiliary known (for many auxiliaries it is often unclear whether omission has occurred, and if so, what form the auxiliary should take). Omission of auxiliary BE can be identified by the presence of the progressive verb inflection while omission of auxiliary HAVE can be identified by the presence of a perfect verb form. Second, we evaluate four possible explanations for auxiliary omission. The first three explanations attempt to explain auxiliary omission within a generativist framework in terms of single mechanisms operating on an abstract grammar. The fourth approach assumes a multifactor, usage-based framework for acquisition, and focuses specifically on the role of input frequencies.

The results show that none of the single mechanism accounts situated within a generativist framework can adequately account for the data. In contrast, the frequency patterns in the input appear to play an important role in early auxiliary use and nonuse, and where frequency does not predict the pattern of acquisition, there are principled explanations for these findings that are consistent with a constructivist framework.

1.3. *A generative grammar interacting with a lack of lexical knowledge*

The first possible explanation for auxiliary omission that must be ruled out before other explanations can be considered is that children's early omission simply reflects a lack of lexical knowledge. Auxiliaries BE and HAVE take different lexical forms according to the person, number and tense of the sentence (BE: *am, is, are, was, were*; HAVE: *have, has, had*). Therefore, it is possible that patterns of omission reflect children's lack of knowledge of the required lexical form. The apparently alternating use of sentences containing auxiliaries and those where the auxiliary is omitted could reflect the consistent use of one form of the auxiliary, and the omission of a different form that children have not yet acquired. If children's early auxiliary use reflects an abstract underlying grammar, and a lack of lexical knowledge is responsible for the patterns of auxiliary omission observed in children's early speech, we would predict that although children's overall levels of auxiliary omission may be high, auxiliary omission in children's speech will be low for individual auxiliary forms once they are acquired.

1.4. *A generative grammar interacting with performance limitations*

A second possible explanation for early auxiliary omission is that omission reflects performance limitations in production (Valian 1991). From this perspective, children are thought to operate with an abstract underlying grammar but their ability to overtly realize their knowledge depends on the performance demands of the intended utterance. Valian argues that the production of modals and auxiliaries without a subject argument (e.g., *can go, is eating*) is unacceptable in the adult grammar, and provides evidence that, she suggests, shows English-speaking children are sensitive to this property of the adult language. As the performance demands in producing grammatical sentences including both a sentence subject and an auxiliary are relatively high, especially at the early stages of development, performance limitations in production might therefore result in auxiliary omission.

Valian (1991) put forward a similar argument for early subject omission, and measured the effects of performance limitations in production

by calculating the mean length of utterance (MLU) of verb phrases (VPs) that appear with a subject argument and comparing this to the MLU of VPs where the subject is omitted (see also Bloom 1990). She suggested that if producing a subject has an associated processing cost, children will have greater processing resources available when the subject is omitted and, therefore, subjectless VPs should be longer than VPs produced with a subject. Similar predictions can be derived with respect to early auxiliary omission. If producing an auxiliary has an associated processing cost, sentences without an auxiliary should be significantly longer than utterances with an auxiliary (when the auxiliary is omitted from the calculation following Valian's methodology). If there is no difference in the length of sentences where the auxiliary is omitted and those where the auxiliary is produced (excluding the auxiliary), this would suggest that producing an auxiliary does not carry a cost in terms of sentence production, and therefore that auxiliary omission does not reflect performance limitations in production.

1.5. *A generative grammar incorporating the Optional Infinitive hypothesis*

A third possible explanation for auxiliary omission is provided by the Optional Infinitive (OI) hypothesis (Wexler 1994). Wexler claims that children initially believe that tense and agreement marking is optional. Auxiliaries BE and HAVE carry both tense and agreement, and therefore optional marking of these grammatical features would result in the production of a number of utterances with and without auxiliaries. Thus, this model suggests that although children may have acquired the necessary auxiliary forms, they will continue to omit auxiliaries because they lack the knowledge that tense and agreement marking is obligatory. Researchers working within the Optional Infinitive framework have claimed that the proportional correct use of finite forms in obligatory contexts increases over time, and in a similar manner for a number of different markers of tense (regular and irregular past tense, the third-person singular, BE and DO) (Rice et al. 1995; 1998; 2000). They take this as evidence that children's early use of finite forms, including auxiliaries, reflects an unitary phenomenon that underlies the use of all forms marking tense and agreement, and that matures over the course of development. One prediction that follows from this model is that once children have acquired the necessary lexical forms, they should show similar rates of omission with different forms of the same auxiliary. If differences are found in the levels of auxiliary provision between auxiliary forms, this would suggest that system-general mechanisms operating at the abstract level of

tense and agreement may not provide the best explanation for early auxiliary omission.

1.6. *A usage-based approach incorporating the role of input frequency*

1.6.1. *The relation between children's lexically based grammars and input frequency* The three explanations for auxiliary omission outlined above share the assumption that a single mechanism (lack of lexical knowledge, performance limitations, optional marking of tense and agreement) can account for the patterns of auxiliary use and nonuse observed in children's early speech. In contrast, constructivist models of language acquisition view children's early language use as much more critically dependent on the precise distribution and frequencies of forms in the input and how this interacts with children's current knowledge of lexically based or more abstract constructions, their sociocognitive interests, and the semantic complexity, functional transparency, and phonological salience of the forms. This approach is supported by recent research that suggests that children's early grammatical knowledge may initially be tied to particular lexical items and only gradually becomes more complex and adultlike (Tomasello 1992; Pine et al. 1998). Furthermore, children's knowledge of individual lexically based frames and the development of more abstract linguistic schemas is thought to reflect the specific properties of the language they hear. Recent research suggests that relationships exist between children's early language use and the language they hear when examined at the lexical level, providing support for this approach (Rowland and Pine 2000; Theakston et al. 2001, 2002).

Previous research on the role of the input in the acquisition of the auxiliary system has implicated maternal questions of different types in the learning of the auxiliary system, but there is little agreement as to which question types in particular are important for acquisition (see Richards 1994 for an overview). However, if children's early knowledge of syntax is tied to individual lexical items, this suggests that to determine relations between the language children hear and their acquisition of auxiliaries, we should examine the input with respect to the individual lexical forms available to children, without making assumptions about the abstract status of their grammars. In the following section, we focus on the role of input frequencies in a constructivist, usage-based approach to children's use and nonuse of auxiliaries.

1.6.2. *The role of input frequency in auxiliary use and the development of abstract constructions in children's speech* One factor that might be expected to affect children's auxiliary use and nonuse is the frequency of

individual subject + auxiliary combinations present in the language children hear. If children begin the language-learning process by acquiring a small number of lexically specific constructions that are based on high-frequency combinations that appear in the speech they hear (Pine et al. 1998; Tomasello 1992, 2000; Theakston et al. 2001), with respect to auxiliary provision, this might mean that they begin with a series of independent constructions, for example, *I'm V-ing*, *He's V-ing*. Over the course of development, children are assumed to induce increasingly abstract constructions by developing variable slots in previously lexically specific frames (Peters 1983; Lieven et al. 1997), thus they may only gradually develop the more abstract construction *X'BE V-ing*. There is likely to be a complex interaction between input frequency of various types, and the extent to which children rely on lexically specific versus more abstract constructions in their production of utterances with and without an auxiliary. Even studies of adult language use suggest that some lexical phrases such as *I don't know* may be stored as linguistic wholes that coexist with more abstract schemas such as *I don't V*, or pronoun *N don't V*, due to their high token frequency in adult language (Bybee and Scheibman 1999). In contrast, linguistic schemas that are of high type frequency will be more productive, i.e., more likely to support generalization to new linguistic items (Bybee 1997). From this perspective, the most important aspects of the input in determining the particular linguistic items acquired early by children are thought to be the type and token frequencies of particular lexical items and constructions in the input (Croft 2001; Langacker 1987).

By definition, constructions are paired representations of form and meaning, therefore the development of fully abstract constructions is thought to depend on children's growing awareness of both the structural and semantic similarities between exemplars (see, e.g., Dabrowska 2000; Israel et al. 2000). Although type and token frequencies in the input are seen as central to the development of children's linguistic representations, exactly how children's lexically based constructions develop into the more abstract schemas that underlie adult language use is not fully understood. Abstraction may be at least partially driven by children gaining a fuller understanding of the semantics of more lexically based constructions. One possibility is that the process of schema abstraction in language acquisition shares commonalities with the development of abstraction in other areas of cognition such as categorization, where categories may be represented as networks of more closely and loosely connected exemplars organized around prototypes (Goldberg 1995, Munakata et al. 1997). From a constructivist perspective, the acquisition of linguistic knowledge is not an all-or-nothing phenomenon. Children are initially expected to have only a partial knowledge of individual linguistic constructions,

or to operate with weakly specified constructions that gradually gain a stronger representation in the linguistic system through increasing linguistic experience. According to some researchers, changes in the strength and/or connectedness of children's linguistic representations might explain the differences between the results of preferential looking, comprehension, and act-out and production studies investigating children's early syntactic knowledge (Tomasello and Abbot-Smith 2002).¹

Although very little is known about the kinds of neural structures or probabilistic networks that might underlie language representations, it is possible to derive a number of theoretically motivated predictions concerning children's early language use. First, the more frequently a specific construction is modeled in the input, the more likely children are to learn the construction as a whole, and therefore produce the construction in their own speech. Second, constructions that are modeled less frequently in the input, or that appear with greater variability in their constituent parts are likely to be more difficult for children to learn, will initially have a weaker or partial representation in children's linguistic systems, and will consequently be less likely to be produced in children's speech. With respect to the acquisition of auxiliary syntax, we might expect that children will produce a specific auxiliary if they have acquired a construction containing that auxiliary based on frequent use in the input. If children have not yet acquired an appropriate construction, or have only a weakly or partially represented construction, they may rely on an alternative construction that has a stronger representation in their linguistic systems. Alternatively, Tomasello (2000) suggests that at all stages of development, children use a process of structure combining to produce novel utterances (i.e., utterances not learned directly from the input) by combining previously acquired constructions. In the case of auxiliary omission, children may combine previously acquired lexical subjects with previously acquired verbs, or substitute different verbs in a lexically based construction to create novel utterances that do not contain an auxiliary (e.g., *I + V*). Clearly, these processes rely on children having some kind of abstract knowledge of subject-predicate constructions in that their early word combinations are motivated by both pragmatic or semantic considerations and typically conform to the word order of the target language. However, it is important to note that the kind of knowledge thought to underpin these rather limited instances of productivity is very different from the kinds of abstract knowledge assumed under generativist approaches to grammar. The circumstances under which children will produce a weakly or partially represented construction are not yet understood, but could involve probabilistic competition between alternative structures (for instance, *I'm going* versus *I going*) and/or a degree of

priming such that constructions that have recently been produced by the child or modeled in the input may be more likely to be produced than a weak or partially represented construction that has not recently been activated in some way.

Thus, if there is an interaction between the lexically based nature of children's early speech and the properties of the input that influences early auxiliary use and nonuse, two predictions can be made:

1. If children's early knowledge of auxiliary syntax is lexically specific and closely based on the language they hear, the order in which children acquire specific subject + auxiliary combinations, for example *He's V-ing*, will reflect their relative frequencies in the language children hear.
2. There is likely to be an interaction between type and token frequency that affects children's proportional use of different auxiliary forms.
 - a. High-frequency subject + auxiliary combinations that are fully lexically specified and therefore have high token frequency (e.g., *He's*), are likely to be learned early and stored as linguistic wholes. These forms should therefore have a strong representation in the children's linguistic systems and be produced frequently, leading to high rates of auxiliary provision for these forms.
 - b. Sequences where there is more variation in subject form, for example NP's, will require children to develop a more abstract schema on which to base correct auxiliary use. As the development of abstract schemas is assumed to take place gradually, schemas with variable slots will initially have weaker representations in the children's linguistic systems than those that are completely lexically specified. Therefore, auxiliary provision for these subject forms is likely to be lower in obligatory contexts than for fixed subject + auxiliary combinations.
 - c. Low-frequency combinations in the input are likely to be acquired later in development, and will also be weakly or partially represented in the children's linguistic systems when initially acquired due to their low frequency of use in the input. Auxiliary provision is therefore likely to be initially low for these forms.

1.7. *Aims*

The purpose of the present article is to examine patterns of auxiliary omission for auxiliaries BE and HAVE in a longitudinal dataset from twelve children followed between the ages of two and three years. The data will be examined to evaluate the different explanations outlined above for auxiliary omission, and to attempt to determine how early

auxiliary use can inform our understanding of the underlying system available to children during the early stages of language acquisition.

2. Method

2.1. *Subjects*

The children in this study were eleven of twelve children who took part in a longitudinal study of early language development. One child only produced three tokens of auxiliary BE and no tokens of auxiliary HAVE and was excluded from the analysis. The children were from predominantly middle-class families and were recruited through newspaper advertisements and local nurseries. They were first-borns from monolingual English-speaking families, and were cared for primarily by their mothers. At the beginning of the study the children ranged in age from 1;8.22 to 2;0.25 with a mean length of utterance (MLUs) between 1.06 to 2.22 in morphemes ($\bar{x} = 1.60$, $SD = 0.39$).

2.2. *Procedure*

The children were audio-taped in their homes for an hour on two separate occasions in every three week period for one year. They engaged in normal everyday interaction with their mothers. The data were orthographically transcribed using the CHILDES system (MacWhinney and Snow 1990; MacWhinney 2000), and are available on the CHILDES database (The Manchester corpus, Theakston et al. 2001).

2.3. *The children's corpora*

Using the CLAN programs (MacWhinney 2000), the children's data were searched for all instances of progressive verbs produced in declaratives for auxiliary BE, and all instances of perfective verbs produced in declaratives for auxiliary HAVE.² Self-repetitions and imitations, incomplete utterances, partially intelligible utterances, and routines (counting, nursery rhymes, etc.) were excluded from the children's corpora. Only those utterances including a subject argument were included in the analysis, as these are the only utterances where (a) auxiliary omission can be reliably identified, and (b) the form of the omitted auxiliary can be accurately determined. All the children's utterances were coded for the presence or absence of the required auxiliary and the proportional use of each auxiliary calculated for each three-week period. The order of emergence of each form of auxiliary BE and each form of auxiliary HAVE was recorded for each child.

2.4. *The mothers' speech corpora*

Each mother's data from the first eight tapes were searched for all exemplars of auxiliary BE and auxiliary HAVE. Only utterances including both an auxiliary and a subject argument were included in the analysis.

3. Results

3.1. *The overall pattern of auxiliary use*

Table 1 shows the number of auxiliary contexts and the mean level of auxiliary provision, based on three-week periods, for BE and HAVE for each child across the whole year. Overall, these children omit auxiliaries approximately 65 percent of the time. Figures 1 and 2 show the levels of provision of BE and HAVE (respectively) for each child at the beginning of the study, approximately six months into the study, and at the end of the study. Although auxiliary provision increases throughout the study, the children still omit the auxiliaries approximately 40 percent of the time at the end of the study.³

We therefore have to account for the pattern of omission that occurs after the children have begun to produce these auxiliaries and therefore, in principle, have the necessary grammatical knowledge required to support correct auxiliary use.

Table 1. *Mean auxiliary provision based on three-week periods for the year*

	BE			HAVE		
	Mean provision (%)	SD	Number of utterances	Mean provision (%)	SD	Number of utterances
Anne	53.8	21.1	427	50.2	27.3	216
Aran	30.7	28.3	260	26.3	24.6	633
Becky	43.8	30.3	512	42.0	26.1	338
Carl	39.1	25.1	914	36.2	23.2	389
Dominic	15.7	16.5	336	11.1	14.6	329
Gail	53.8	18.0	389	60.0	23.9	248
Joel	47.6	33.6	313	31.9	25.3	359
John	38.5	22.8	144	25.8	25.1	131
Liz	28.8	23.7	435	16.0	13.9	285
Nicole	8.5	10.3	209	15.4	19.7	105
Warren	35.6	21.5	566	39.6	21.8	358
<i>Mean</i>	36.0	22.8	409.5	32.2	22.3	308.3

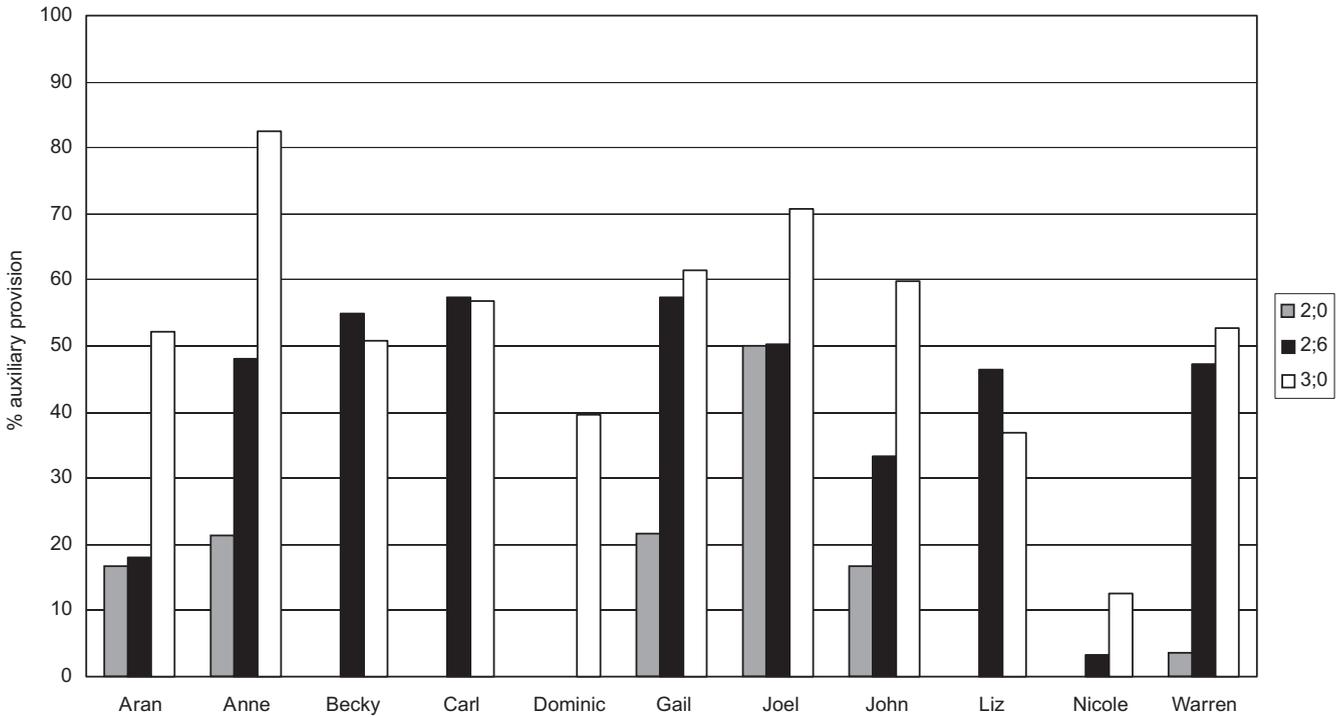


Figure 1. Proportional auxiliary provision with BE at approximately 2;0, 2;6, and 3;0 for each child

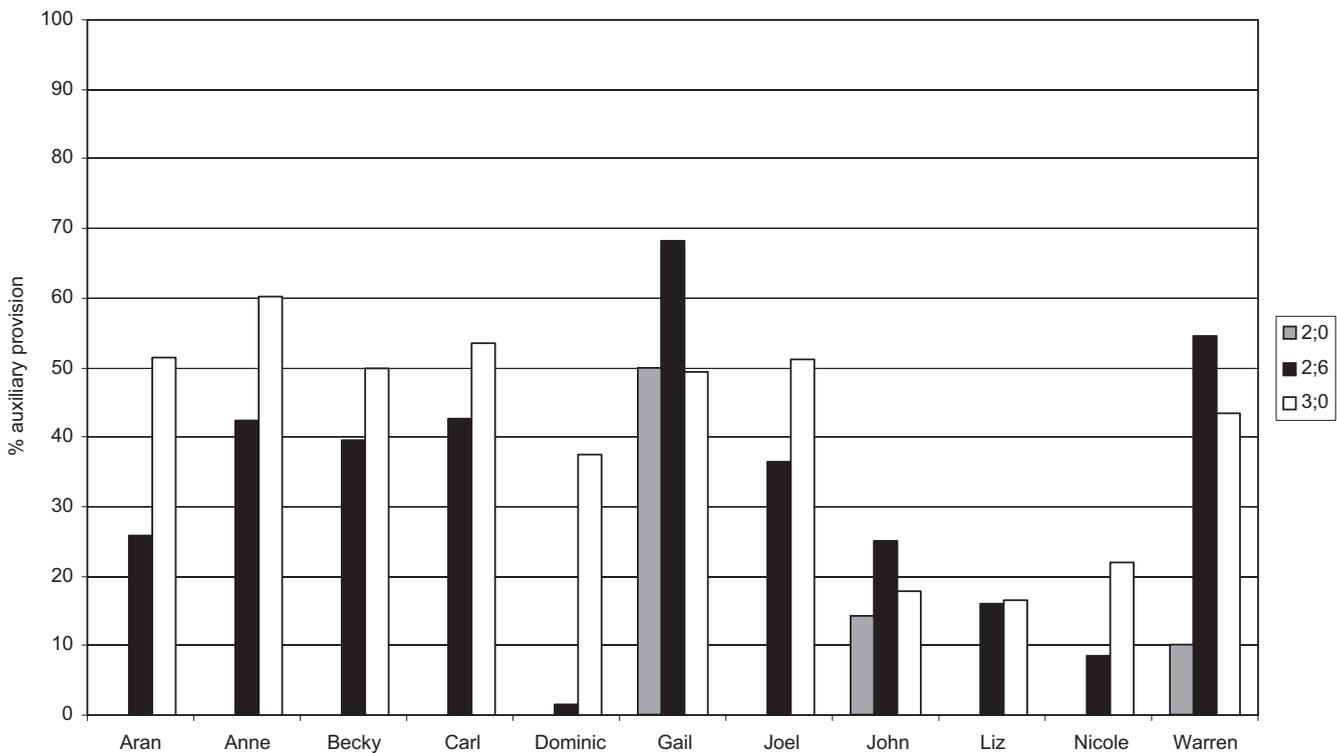


Figure 2. Proportional auxiliary provision with HAVE at approximately 2;0, 2;6, and 3;0 for each child

3.2. *Knowledge of individual lexical forms*

To determine whether a lack of knowledge of the necessary lexical forms was responsible for auxiliary omission, we calculated the levels of provision of the earliest acquired and most frequently produced auxiliary forms from the point in development at which they were acquired. For this analysis, acquisition was defined as the second use of a particular auxiliary form in an individual child's speech. The most common and earliest acquired forms of BE produced by the children were contracted *is* and contracted *am*; the most common and earliest acquired forms of HAVE produced by the children were contracted *has* and contracted *have*.⁴ Combining the data for *is* and *am*, and the data for *has* and *have*, once the forms are acquired the mean levels of auxiliary provision were 50.4 percent for BE (SD = 21.3) and 51.2 percent for HAVE (SD = 22.4). This shows that even after the children have acquired the necessary auxiliary forms, they continue to omit auxiliaries, suggesting that a lack of lexical knowledge cannot account for auxiliary omission.

3.3. *A performance limitations account*

To determine whether performance limitations are responsible for auxiliary omission, data from three different time periods were selected for analysis. As performance limitations would be expected to decrease over the course of development, the largest effects should be observed at the earliest stages of auxiliary use. However, as these children were still omitting auxiliaries at the end of the study, if performance limitations are to fully explain auxiliary omission, they must be able to explain omission at all stages of development. Data from four tapes when each child first produced a form of the auxiliary in question, four tapes taken approximately six months into the study, and the last four tapes were examined. For auxiliaries BE and HAVE, the mean length of utterance (MLU) of utterances with and without the auxiliary was calculated at each time period. If performance constraints are responsible for auxiliary omission, those utterances where the auxiliary is omitted should be significantly longer than those where the auxiliary is produced (when the auxiliary is excluded from the calculations). For both BE and HAVE, *t*-tests revealed that there were no significant differences in length between utterances containing the auxiliary (excluding the auxiliary constituent) and those where the auxiliary was omitted at any stage of development (nonsignificant *t* values ranged from -2.10 to 2.90). This suggests that performance limitations are not playing a central role in auxiliary omission, at least using mean length of utterance (MLU) as a measure of limitations.

3.4. The Optional Infinitive hypothesis

The Optional Infinitive hypothesis predicts that the rate of auxiliary provision should be similar across different auxiliaries and different forms of those auxiliaries, reflecting an underlying abstract system governing tense and agreement marking. First, the data were examined to determine whether the overall pattern of omission with auxiliary BE and auxiliary HAVE was similar for individual children. A correlation comparing the children's mean auxiliary provision with BE with their mean auxiliary provision with HAVE across the whole year (see Table 1) was significant ($r = 0.87, df = 9, p < 0.01$), thus providing support for the Optional Infinitive hypothesis at the level of overall provision. However, this analysis does not take into account the stage at which individual children first begin to produce specific auxiliary forms. It is possible, therefore, that the overall figures mask differences in the children's levels of provision of individual forms of the two auxiliaries.

The data were examined further to determine whether, once the children had acquired particular auxiliary forms, they demonstrated the same degree of optionality with each form of the auxiliary in question. At the level of individual auxiliary forms, the children's levels of provision of *am* were compared with their levels of provision of *is*, and their levels of provision of *has* were compared with their levels of provision of *have*. The data show that there are differences in the children's levels of provision of individual forms of BE and HAVE. For auxiliary BE, the children provide the form *is* significantly more frequently than the form *am* in obligatory contexts (*is*: $\bar{x} = 61.9\%$, *am*: $\bar{x} = 32.7\%$; $t = 3.52, df = 9, p < 0.01$), while for auxiliary HAVE, they provide the form *has* significantly more frequently than the form *have* in obligatory contexts (*has*: $\bar{x} = 62.4\%$, *have*: $\bar{x} = 25.5\%$; $t = 3.89, df = 7, p < 0.01$). This suggests that rather than the children's auxiliary use being determined by an underlying optionality that would be expected to apply equally across different forms of the same auxiliary, these children have different knowledge regarding the behavior of particular auxiliary forms. This is most markedly the case for the form *have*, typically acquired somewhat later in development than the form *has* and which therefore, if optionality decreases as the child matures, should be less susceptible to optional omission than other, earlier learned auxiliary forms.

A further analysis was then carried out to determine whether the children's use of specific forms of BE and HAVE was consistent across different types of lexical subject, or whether in fact differences in the level of auxiliary provision can also be observed for an individual auxiliary form, dependent on the specific subject produced. The children's data

Table 2. *Auxiliary BE: Proportional auxiliary provision with individual subject + auxiliary combinations once acquired (acquisition defined as the second use of a given combination) for forms with over five obligatory contexts after acquisition*

	NP's	Proper noun's	<i>he's</i>	<i>it's</i>	<i>she's</i>	<i>that's</i>	<i>I'm</i>	<i>you're</i>	<i>we're</i>	<i>they're</i>
Anne	39.1	54.2	76.9	80.6	66.7		57.5		50.0	71.4
Aran	86.7	42.9	79.4	83.3			20.4			
Becky	55.0	40.0	90.5	97.4	87.5		61.2	40.0	27.3	29.2
Carl	44.4	43.9	84.0	92.8			22.6			33.3
Dom	19.2	11.1	38.5	50.0			26.1	41.7		
Gail	77.1	63.5	96.9	100			29.7	33.3	66.7	
Joel	66.7	88.9	91.7	100			66.0			37.5
John	28.1		100	84.0						
Liz		45.5	80.0	83.3			25.0		16.7	33.3
Nicole		1.4		80.0			16.0			
Warren	53.0	42.9	75.8	80.5		50.0	5.4			
Mean	52.1	43.4	81.4	84.7	77.1	50.0	33.0	38.3	40.2	40.9

were coded for the following high-frequency subject + auxiliary combinations (BE: *he's*, *she's*, *it's*, *that's*, *I'm*, *you're*, *we're*, *they're*; HAVE: *he's*, *she's*, *it's*, *that's*, *I've*, *you've*, *we've*, *they've*), and the more general high-frequency syntactic categories (NPs and Proper nouns for BE and HAVE). The children's levels of auxiliary provision were calculated for each subject from the point at which they had produced two instances of the specific subject + auxiliary combination to determine whether there were differences in levels of provision of individual auxiliary forms with different subject + auxiliary combinations. Tables 2 and 3 show the children's proportional auxiliary use with specific subject + auxiliary combinations for auxiliaries BE and HAVE, respectively.

It is clear that there is considerable variation in the levels of auxiliary provision with different lexical subjects. One-sample *t*-tests were carried out to determine whether the levels of provision of individual subject + auxiliary combinations differed significantly from the children's mean levels of provision of BE and HAVE. For these analyses, the overall level of provision of contracted *is*, *am*, and *are*, and the overall level of provision of contracted *have* and *has*, were calculated for each child, based on the data for each form from the point at which s/he produced two examples of that contracted form, and the mean level of provision across children was calculated. The results show that for auxiliary BE, the combinations *he's* and *it's* were produced in obligatory contexts significantly more frequently than the mean level of auxiliary BE provision (*he's*: $t = 6.02$, $df = 9$, $p < 0.001$; *it's*: $t = 8.58$, $df = 10$, $p < 0.001$),

Table 3. Auxiliary HAVE: Proportional auxiliary provision with individual subject + auxiliary combinations once acquired (acquisition defined as the second use of a given combination) for forms with over five obligatory contexts after acquisition

	NP's	Proper noun's	<i>he's</i>	<i>it's</i>	<i>she's</i>	<i>that's</i>	<i>I've</i>	<i>you've</i>	<i>we've</i>	<i>they've</i>
Anne	53.8	78.3	87.5	85.7	50		32.8		33.3	
Aran		83.3	83.3	71.6						
Becky	66.7	75.0	90.9	88.5	83.3		13.0	30.0	22.7	
Carl	52.2	60.0	81.3	93.3						
Dom	16.7		60.0	43.8			8.6	45.5		
Gail	90.9	73.0	91.7	89.8	100		22.7			
Joel	57.1		94.4	100			18.5	22.7	33.3	
John		20.0		88.6						
Liz	50.0		50.0	90.0						
Nicole		5.0								
Warren	48.4	46.2	75.0	85.9		66.7				
Mean	54.5	55.1	79.3	83.7	77.8	66.7	19.1	32.7	29.8	

while the combination *I'm* was produced significantly less frequently than the mean level of provision for BE ($t = -2.35$, $df = 9$, $p < 0.05$). For auxiliary HAVE, the combinations *he's* and *it's* were produced significantly more frequently than the mean level of provision for HAVE (*he's*: $t = 5.55$, $df = 8$, $p < 0.01$; *it's*: $t = 6.53$, $df = 9$, $p < 0.001$), while the combinations *I've* and *we've* were produced significantly less frequently than the mean level of provision for HAVE (*I've*: $t = -7.69$, $df = 4$, $p < 0.05$; *we've*: $t = -6.07$, $df = 2$, $p < 0.05$). Thus, although there are some similarities in the level of auxiliary provision between individual subject + auxiliary combinations, differences exist, both between auxiliary forms (e.g., *he's* versus *I'm*) and between combinations with specific auxiliary forms (e.g., *he's* versus NP's), that suggest that the children's auxiliary use and nonuse is unlikely to reflect the workings of an underlying abstract system of tense and agreement marking.

3.5. The role of input frequencies

If children learn how to use auxiliaries by beginning with specific lexical items, and only gradually building up a more abstract representation of auxiliary syntax, we would expect that the specific subject + auxiliary combinations children acquire earliest will be those that appear with the greatest frequency in the language they hear. The prediction, therefore, is that the frequency of use of individual subject + auxiliary combinations in the input will predict the relative age at which they are acquired by

the children.⁵ To test this prediction, the mean frequencies of use of individual subject + auxiliary combinations (see Tables 2 and 3) in the mothers' speech were compared with their mean age of acquisition, calculated in days, in the children's speech. Combinations that were not acquired by individual children during the course of the study were assigned an age of acquisition seven days after the last recording for that child. For both auxiliary BE and auxiliary HAVE, the correlation between mean frequency in the mothers' speech and mean age of acquisition in the children's speech was nonsignificant (BE: $r = -.04$, $df = 8$, $p > 0.05$; HAVE: $r = -.04$, $df = 8$, $p > 0.05$). However, a striking observation is the fact that although by far the most frequent subject + auxiliary combinations present in the input are the forms *you're* and *you've* for auxiliaries BE and HAVE respectively (these forms are between approximately twice and thirty-four times more frequent than the other combinations), these forms are typically acquired late, and are only acquired by three of the children during the course of the study. Clearly, these forms depend on something in addition to input frequency for their acquisition. When these second-person pronouns were excluded, there was a significant correlation between the mean frequency of individual subject + auxiliary combinations in the input, and their mean age of acquisition in the children's speech for both auxiliaries (BE: $r = -.71$, $df = 7$, $p < 0.05$; HAVE: $r = -.70$, $df = 7$, $p < 0.05$). This suggests that with the exception of the pronominal form *you*, the frequency of use of specific subject + auxiliary combinations in the input is an important determiner of their age of acquisition in the children's speech.

The data were then analyzed further to determine whether the predictions generated from a lexical-learning account were supported. More specifically, although the frequency of use of specific subject–auxiliary combinations in the input predicts their age of acquisition in the children's speech, it is unclear whether the token frequency of individual combinations and the type frequency of the lexical subject (fixed versus variable) in different combinations are related to the children's proportional provision of auxiliaries with individual subjects. Those subject–auxiliary combinations acquired by five or more of the children were examined further. For auxiliary BE, these forms were NP's, Proper noun's, *he's*, *it's*, *they're*, and *I'm*; for HAVE these forms were NP's, Proper noun's, *he's*, *it's*, and *I've*. Table 4 shows the forms for BE and HAVE listed in order of their mean age of acquisition, alongside the children's mean levels of auxiliary provision for each form, and an indication of the relative frequency of use of each form in the input. Specific combinations were coded as high frequency if there was a mean of ten or more exemplars in the input, and low frequency if there was a mean of nine exemplars or less

Table 4. Order of acquisition and mean levels of auxiliary provision for subject + auxiliary combinations acquired by more than five children

Order acquired	BE			HAVE		
	Subject + auxiliary combination	Frequency in input ^a	Auxiliary provision (%)	Subject + auxiliary combination	Frequency in input	Auxiliary provision (%)
1	<i>I'm</i>	H	33.0	<i>it's</i>	H	83.7
2	<i>it's</i>	H	84.7	Proper noun's	H	55.1
3	Proper noun's	H	43.4	<i>he's</i>	H	79.3
4	<i>he's</i>	H	81.4	NPs	H	54.5
5	NPs	H	52.1	<i>I've</i>	H	19.1
6	<i>they're</i>	L	40.9			

^a Combinations were coded as high frequency (H) if there was a mean frequency of ten or more exemplars, and low frequency (L) if there was a mean frequency of nine exemplars or less in the input.

in the input. For this analysis, the general categories of NP's and Proper noun's were examined further to determine (a) whether there were any individual combinations (e.g., *Mummy's*, *child's name*) that met the criteria for high-frequency lexical combinations, and (b) whether there was a sufficiently high type frequency in these categories to justify coding the data with respect to the broad categories of NP and Proper noun rather than in terms of individual subject + auxiliary combinations. Results revealed that no individual NP or Proper-noun subjects occurred with high frequency (a mean of ten or more exemplars) in the input. In addition, there was a mean of 13.6 different NP subjects with BE (\bar{x} = 13.2 with HAVE) and 5.6 different Proper-noun subjects with BE (\bar{x} = 6.5 with HAVE), showing that these combinations occur with relatively high type frequency in the language children hear. This suggests that children cannot rely exclusively on lexically based learning, and must acquire a relatively abstract schema to support auxiliary use with these types of subjects.

For BE, there are five high-frequency combinations, and one low-frequency combination acquired by five or more of the children. It is clear that among the high-frequency forms, those with fixed lexical subjects (*he's*, *it's*) appear with much higher levels of auxiliary provision than those with variable subjects (NP's, Proper noun's). This pattern of results is predicted by a lexical learning account that suggests that frames with fixed lexical subjects will be easier for children to acquire, and have stronger representations in children's linguistic systems, than those where a process of abstraction is required to generate a frame available for use with a range of different subject forms. Moreover, the form *they're* is infrequent in the input. Therefore, although the subject is fixed, this form is predicted to have a weaker representation in the children's linguistic systems when first acquired, leading to initially lower rates of auxiliary provision than higher frequency combinations with fixed lexical subjects. The children's data provide support for this prediction. However, their use of the first-person form *I'm* does not follow the predictions derived earlier. This form has high frequency in the input, is acquired early, and yet has relatively low levels of auxiliary provision in the children's speech.

For HAVE, there are five subject + auxiliary combinations acquired by five or more of the children, and all are used with high frequency in the input. However, it is again possible to observe the predicted pattern of results with the exception of the first-person form *I've*. The two combinations with fixed subjects (*he's*, *it's*) have a much higher rate of auxiliary provision than the two combinations with variable subjects (NP's, Proper noun's). Again, the first person form *I've* is used with high frequency in the input, is acquired relatively early in development, but displays much lower levels of auxiliary provision in the children's speech than is

predicted by a lexical learning account. It is interesting to note that the subject + auxiliary combinations, *he's* and *it's*, that are used with both BE and HAVE occur with high rates of auxiliary provision, while the forms *I'm* and *I've* that encode only one of the auxiliaries have lower rates of auxiliary provision. It seems likely that the overall frequency in the input of those forms that are shared across BE and HAVE contributes towards increased levels of auxiliary provision. Although children are thought to be gradually building up independent representations of the meanings associated with the progressive and present perfect constructions, it is likely that the units *he's* and *it's* are also represented individually. This means that any use of these forms in the input will strengthen the representation of these forms independently of the specific constructions in which they appear, and therefore the likelihood that children will retrieve the forms in their own productions, leading to increased auxiliary provision, will increase (copula constructions, e.g., *he's sad*, could also impact on the role of input frequency in the acquisition of auxiliary forms).

If we look back to Tables 2 and 3, we can see the rates of auxiliary provision for those forms acquired by fewer than five of the children. Although we must be tentative in the conclusions we draw based on these limited data, it appears that for both BE and HAVE, the rates of provision with these forms follow the predictions of a constructivist account. For both BE and HAVE, the forms *she's* and *that's* are produced with relatively high levels of auxiliary provision. Although these forms are acquired relatively late in development, they are (a) fixed subject forms, and (b) fit an existing schema in the children's linguistic systems acquired with other, earlier learned third-person forms (X's). In contrast, the forms *you're*, *we're*, *you've*, and *we've* are produced with relatively low levels of auxiliary provision. Although these forms also have fixed subjects, they do not conform to any larger pattern that could form the basis for a schema that the children might use, and they must therefore learn each form independently. Thus, the pattern of results is broadly consistent with a constructivist account where an interaction between input frequency, the consistency of the lexical subject, and the gradual emergence of abstract schemas determines levels of auxiliary provision in early child speech.

4. Discussion

This study examined a number of possible explanations for auxiliary omission with BE and HAVE in children's early language to determine which explanation most closely matched the data. First, the possibility

that the children had not yet acquired the required lexical forms of the auxiliaries was examined. The data show that even after particular auxiliary forms were acquired, the children continued to frequently omit those forms in obligatory contexts. This suggests that it was not a lack of lexical knowledge that determined the observed pattern of auxiliary omission.

Second, the data were examined to determine whether performance limitations in production might account for the pattern of early omission. Analysis showed that when measured in terms of mean length of utterance at three different stages of development, the children's utterances where the auxiliary was omitted were not significantly longer than those where the auxiliary was produced, suggesting that performance limitations were not responsible for early auxiliary omission.

Third, the suggestion that children initially omit auxiliaries on some occasions but not others due to an immature understanding of the obligatoriness of tense and agreement marking (the Optional Infinitive hypothesis, Wexler 1994) was considered. If the degree to which children mark tense and agreement in obligatory contexts is determined by system-general maturational constraints, we would expect that different forms of the same auxiliary, and different lexical subjects will be similarly affected with respect to the degree of optionality of the auxiliary. However, the data show that there were significant differences between both different forms of the same auxiliary, and individual lexical subjects with respect to the relative provision of auxiliaries. This suggests that the children may have different knowledge of the syntactic behavior of individual auxiliary forms, rather than auxiliary production being governed by innate maturational constraints. Thus, although it is possible to describe the overall pattern of auxiliary omission in terms of the optional marking of tense and agreement in early child speech, it is clear that this model fails to provide any precise predictions regarding the exact pattern of omission we would expect to find with respect to individual lexical items in the children's speech. Further development of the model to incorporate a clearly specified role for performance limitations or lexical preferences is needed if the model is to explain the current data (see also Theakston et al. 2003).

These results suggest that single-factor models are unlikely to provide a satisfactory account of auxiliary use and nonuse in the early stages of language acquisition. Instead, more complex models that take into account a number of competing factors are needed to explain the data, in particular the variation in levels of auxiliary provision across individual lexical subjects. The final analyses represent a first attempt to evaluate a frequency-dependent, usage-based account of auxiliary use and nonuse, taking into account the frequency and distribution of individual lexical forms in the

language children hear and the interaction between the properties of the input and children's lexically based knowledge of grammar.

The relation between lexically based learning and the type and token frequencies of individual lexical combinations in the input was examined. The data show that, first, the children's acquisition of individual subject + auxiliary combinations reflected their relative frequency of use in the language to which they were exposed, with the exception of the pronominal form *you*. Second, there was support for the predictions derived from a constructivist approach to early acquisition. High-frequency, fully specified subject + auxiliary combinations in the input that were acquired early by the children tended to show higher levels of auxiliary provision in the children's speech than lower frequency, later acquired, fully specified forms. Moreover, high-frequency forms in the input with variable subjects that were acquired early by the children tended to show lower levels of auxiliary provision in the children's speech than high-frequency forms with fixed lexical subjects. The clear exception to this pattern was first-person subject + auxiliary forms that were high frequency in the input, acquired relatively early, and contained fixed subjects, but displayed low levels of auxiliary provision in the children's speech. Thus, the frequency-driven learning hypothesis is supported, with the exception of the pronouns *I* and *you*.

The predictions derived from a frequency-dependent usage-based approach are therefore supported in part. There is a relation between the type and token frequencies of individual combinations in the input and the relative use of auxiliaries in declaratives in the children's speech, but the relation between these variables is not straightforward. There are (at least) two important aspects of development that cannot be explained by the properties of the input examined in this study. First, the children acquire the second-person forms *you're* and *you've* much later in development than would be predicted on the basis of their frequency in the language children hear. Second, the children's levels of auxiliary provision with the first-person forms *I'm* and *I've* are much lower than would be predicted on the basis of their frequency in the input and their age of acquisition relative to other subject + auxiliary combinations. It is therefore clear that although the predictions derived from frequency-dependent usage-based accounts go some way towards explaining early auxiliary use and nonuse, more sophisticated models must be developed to fully explain the patterns of language use observed in children's early speech. In particular, there are a number of additional factors, all of which are consistent with the broader constructivist framework, that must be considered when attempting to develop models to explain the acquisition of auxiliary syntax.

First, one obvious difference between those auxiliary forms that are modeled frequently in the input, but are used infrequently in the children's speech, and those where the children's use is more consistent with an input-based account is the person encoded. The third-person forms *is* and *has* can be used in exactly the same way in the children's speech as in the input, whereas first and second-person forms require children to reverse their use from that in the language they hear. Thus there is added cognitive complexity, in that once a particular form has been acquired from the input, children must work out the particular context in which it can be used relative to themselves, which might, in part, explain the children's apparent difficulties with these forms. In addition, in the early stages of language acquisition, children show a strong tendency to talk about their own actions rather than those of others. They might be more motivated to acquire first-person forms than their second-person counterparts, overriding the influence of their relative frequencies in the input. However, while cognitive considerations might explain the early acquisition of first-person forms, and the late acquisition of second-person forms, it is less clear why first-person forms are acquired early, but used with high rates of auxiliary omission for a long period of development.

Second, we may need to adopt a more sophisticated approach to the role of the input in the acquisition process to explain children's high rates of auxiliary omission with first-person pronouns. As observed in the current study, the pronoun *you* is very frequent in speech addressed to children, yet they acquire the first-person form *I* much earlier in development, and use it much more frequently than *you*. This suggests that from relatively early in development, children recognize that these forms can be substituted for each other by observing the distributional similarities between *I* and *you*, enabling them to produce the first-person form *I* in frames modeled with *you* in the input (e.g., from Liz's data—Mother: *What're you doing next?* Child: *pink, orange, I doing orange*). It is therefore possible that the use of *you* in the input may influence the use of *I* in the children's speech. Two possibilities will be considered. First, the use of questions in the input directed towards the child that model the pronominal form *you* may contribute towards the high rates of auxiliary omission with the pronominal form *I* in the children's speech (e.g., *Are you drawing?*, *What're you doing?*). In questions, the subject is directly adjacent to the verb without an intervening auxiliary. Therefore, if children model their utterances with the pronominal form *I* on questions in the input with the pronominal form *you*, this may lead to high rates of auxiliary omission with the pronominal form *I*. It is interesting to note that for auxiliaries BE and HAVE, (a) children hear a much larger number of questions with the subject *you* than with any other subject form listed in

Tables 2 and 3 (BE *you*: \bar{x} = 196.7 questions, other subjects: \bar{x} ranges from 0.91 to 25.0 questions; HAVE *you*: \bar{x} = 39.2 questions, other subjects: \bar{x} ranges from 0.73 to 10.0 questions), and (b) the relative proportional use of questions that model S-V word order in comparison with declaratives that model S-Aux-V word order with the pronominal subject *you* in the input is fairly high (\bar{x} = 75.4% questions with BE; \bar{x} = 40.5% questions with HAVE). This means that overall, the pronominal form *you* is modeled just as frequently in S-V word order as it is in S-Aux-V word order. It is therefore possible that the pattern of distribution in the input may contribute towards the particularly high rates of auxiliary omission observed in the children's use of the pronominal subject *I*. Second, if children substitute the pronominal form *you* modeled in the input with the form *I* in their own speech, there is an added complexity in that the auxiliary form used with *you* differs from that used with *I* (*are* versus *am*). The fact that these auxiliary forms are modeled predominantly in contracted form and may therefore be unanalyzed by the children may increase the likelihood that they will fail to provide an auxiliary with first-person subjects in declaratives if they are basing these utterances partly on the use of questions with second-person subjects in the input. More generally, we suggest that there are likely to be a number of complex distributional factors that determine the likelihood that children will produce auxiliaries with first-person subjects, over and above the basic frequency of use of the first-person subject + auxiliary combination *I'm* in the input.

Obviously, this suggestion requires further empirical investigation to determine the extent to which children's use of first-person pronouns might be affected by the use of second-person forms in the input. However, computational simulations of language learning provide some evidence to suggest that building relations between pronouns may be an important development in early language use. An EPAM (Elementary Perceiver and Memorizer) network sensitive to the distributional properties of the input was trained on a sample of input taken from one of the mothers involved in the current study. In the resulting language network generated by the model, some of the first linguistic items to become connected via generative links, thus allowing the interchange of linguistic items in syntactic structure, were pronouns, in particular the pronouns *I* and *you* (see Jones et al. 2000 for full details of the model). Similarly, Oshima-Takane, Takane, and Shultz (1999) showed that a feed-forward computer-network model could correctly acquire the pronouns *me* and *you* when trained on "overheard" speech, replicating the findings of an earlier naturalistic study showing that young children just beginning to produce pronouns are able to correctly reverse use of these pronouns if

exposed to speech not addressed directly to them (Oshima-Takane 1988). If these models at all resemble the processes that occur in language acquisition, this highlights the need to address the question of whether relationships exist between the input children hear and the speech they produce in a rather more sophisticated manner than has previously been attempted, taking into account children's linguistic knowledge at any particular point in development, the specificities of the input to which they are exposed, and how these might interact.

Third, although the data are consistent with a constructivist approach to early language acquisition, it is necessary to consider how children eventually come to reach adultlike levels of productivity. For example, how do children come to realize that auxiliary provision is obligatory in utterances where they have previously omitted auxiliary forms? One possibility is that children acquire an increasing number of lexical frames including the auxiliary, for example *I'm V-ing*, *He's V-ing*, and this leads to the extraction of higher level schemas, for example S-Aux-V.⁶ The extraction of higher level schemas that require a lexical item between the subject of a sentence and its verb may lead children to attend to previously unnoticed auxiliary forms. Similarly, the acquisition of the full rather than contracted forms of auxiliaries may highlight a previously unnoticed slot in the syntactic structure of particular utterances. However, these suggestions require detailed empirical investigation to determine the relation between different aspects of the auxiliary system at different stages in development.

It is clear from the present study that all of the approaches to early auxiliary omission examined fail to offer a complete account of the data, and therefore researchers working within a range of theoretical frameworks need to develop more precise models of the processes governing early language acquisition. A performance limitations account is not supported by the data, and while the Optional Infinitive hypothesis (Wexler 1994) provides a rough description of the data, it fails to provide any specific predictions regarding the precise pattern of acquisition, nor does it explicitly predict the differences observed in the children's levels of provision with individual auxiliary forms or lexical subjects. In contrast, a constructivist approach has the virtue of making specific predictions about the relative levels of provision of specific forms that can then be tested. Some of these predictions were supported by the current analyses while others were not. By focusing on those specific areas of acquisition where constructivist predictions were not supported, we are able to derive new predictions that can in turn be tested to enable us to better understand the complex process of language acquisition. We suggest that there is a need for all accounts to operate at a more sophisticated level to account

for early acquisition of auxiliary syntax, which will need to include reference to children's cognitive development as well as to the complex distributional patterns that children may attend to in the language they hear. Of course, building models of language acquisition that take into account a number of competing influences on children's early language use is difficult, and makes the process of deriving predictions concerning early language acquisition much more complex.

In summary, much more work is needed to more precisely determine the relative influence of a range of different factors (e.g., type and token input frequency, cognitive complexity, phonological salience) on the acquisition of auxiliary syntax, and to more precisely document the relative order of acquisition of particular auxiliary forms in declaratives and questions before we can begin to understand fully how children acquire the complex auxiliary system. This will require detailed studies of early language use that go beyond the scope of the present study, as the children in this study were only just beginning to produce the full range of auxiliary forms as well as auxiliary-fronted questions and tag questions. However, the current study, although rather tentative in its findings, suggests that there may be relationships between the use of auxiliaries in the input and the patterns of auxiliary omission observed in children's speech at the lexical level that may begin to explain auxiliary omission in children's early speech.

Received 18 July 2002
Revision received 14 June 2004

*Universities of Manchester,
Nottingham, and Liverpool,
Max Planck Institute for
Evolutionary Anthropology, Leipzig*

Notes

- * We would like to thank all the families who took part in the research reported here. Thanks also to two anonymous reviewers for their helpful comments. This research was funded by the Economic and Social Research Council, Grant Numbers R000236393 and R000237911. Anna L. Theakston is at the University of Manchester, Elena V. M. Lieven at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Julian M. Pine at the University of Nottingham, and Caroline F. Rowland at the University of Liverpool. Address for correspondence: Anna L. Theakston, Department of Psychology, University of Manchester, Oxford Road, Manchester M13 9PL, UK; e-mail: <anna.l.theakston@manchester.ac.uk>.
1. Studies employing the preferential-looking paradigm typically suggest that children have a greater degree of abstract syntactic knowledge earlier in development than studies that require children to demonstrate their knowledge through the production of nonce words in a range of syntactic structures (see Fisher 2000).

2. The corpora were transcribed in CHAT format and a line of morphological analysis was added using the MOR program. Due to errors in the MOR program, it was not possible to search exclusively for perfect verbs. Thus, all instances of verbs coded as either past tense or perfect on the MOR line were selected. Irregular verbs that take a different form for the perfect and past tense were included in the analysis. Regular verbs, where the past and perfect forms are synonymous were included only if the linguistic context indicated that the form was perfect (e.g., the mother produced the relevant auxiliary or, in the case of *got*, the child was clearly referring to present possession rather than past tense).
3. Although some children (Joel for BE, Gail for HAVE) show high rates of auxiliary provision at the beginning of the study, these figures are based on very small numbers of utterances (one and three respectively) and reflect the use of apparently rote-learned utterances that include a contracted auxiliary form.
4. Some researchers dismiss the use of contracted auxiliary forms in children's early speech claiming that they are unlikely to represent a genuine segmentation of the auxiliary from its subject, and therefore do not represent true auxiliary use (e.g., Pinker 1984). However, children's use of full forms may initially be unanalyzed and tied to lexical frames, and therefore there is no a priori reason to distinguish between full and contracted forms in early auxiliary use. In fact, the majority of the children's auxiliary tokens were contracted for all forms except *have* (*is*: \bar{x} = 90.6%, SD = 9.3; *am*: \bar{x} = 96.7%, SD = 5.3; *has*: \bar{x} = 92.6%, SD = 8.9; *have*: \bar{x} = 35.0%, SD = 28.1). For most of the children, non-contracted forms were produced only during the later stages of development, were infrequent, and, where they were produced, often appeared in utterance-initial position in questions. Thus, during the period of development in question, it is most informative to examine the children's use of contracted auxiliary forms to determine their knowledge of auxiliary use in declaratives.
5. The vast majority of the mothers' auxiliary tokens in declaratives were contracted (*is*: \bar{x} = 91.4%, SD = 4.1; *am*: \bar{x} = 98.1%, SD = 2.5; *has*: \bar{x} = 92.4%, SD = 5.5; *have*: \bar{x} = 83.8%, SD = 6.5), and therefore the analysis of individual subject + auxiliary combinations in the input includes the vast majority of auxiliary forms present in the language children hear. However, this suggests that ignoring the use of contracted forms in children's early speech essentially ignores most of the available data on auxiliary use in both adult's and children's speech.
6. It is also likely that the extraction of these basic level schemas results from abstraction over a number of separately acquired frames with different verbs based on their frequency in the input, e.g., *I'm going*, *I'm singing*. In fact, analysis at the lexical level shows that for seven of the children, auxiliary provision in the frame *I'm V-ing* is significantly higher with *going* than with all other verbs, ($t = 4.94$, $df = 6$, $p < 0.01$) suggesting that their knowledge of the contracted form of *am* is initially tied to the fixed utterance *I'm going* rather than the more flexible *I'm V-ing* frame. The verb *going* is the most frequent verb used in the frame *I'm V-ing* in the input, accounting for a mean of 48.3 percent (SD = 23.6) of all *I'm V-ing* declaratives.

References

- Bloom, Paul
1990 Subjectless sentences in child language. *Linguistic Inquiry* 21, 491–504.
- Bowerman, Melissa
1990 Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics* 28, 1253–1289.

- Bybee, Joan
1997 Three frequency effects in syntax. *Proceedings of the 23rd Annual Meeting of the Berkeley Linguistics Society* 23, 65–85.
1998 The emergent lexicon. In Gruber, M., D. Higgins, K. Olson, and T. Wysocki (eds.), *Papers from the 34th Annual Meeting of the Chicago Linguistics Society, Part 2: The Panels*. Chicago: Chicago Linguistics Society.
- Bybee, Joan and Joanne Scheibman
1999 The effect on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37, 575–596.
- Chomsky, Noam
1957 *Syntactic Structures*. The Hague: Mouton.
- Croft, William
2001 *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Dabrowska, Ewa
2000 From formula to schema: The acquisition of English questions. *Cognitive Linguistics* 11 (1–2), 83–102.
- Fisher, Cynthia
2000 The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition* 82, 259–278.
- Goldberg, Adele
1995 *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Jones, Gary, Fernand Gobet and Julian Pine
2000 A process model of children's early verb use. In Gleitman, L. R. and A. K. Joshi (eds.), *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, 723–728.
- Israel, Michael, Christopher Johnson, and Patricia Brooks
2000 From states to events: The acquisition of English passive participles. *Cognitive Linguistics* 11 (1–2), 103–129.
- Kay, Paul and Charles Fillmore
1999 Grammatical constructions and linguistic generalizations: the “What's X doing Y” construction. *Language* 75, 1–33.
- Lakoff, George
1987 *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Langacker, Ronald
1987 *Foundations of Cognitive Grammar*, vol. 1. Stanford: Stanford University Press.
1991 *Foundations of Cognitive Grammar*, vol. 2. Stanford: Stanford University Press.
- Lieven, Elena, Julian Pine, and Gillian Baldwin
1997 Positional learning and early grammatical development. *Journal of Child Language* 24, 187–219.
- MacWhinney, Brian
2000 *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Erlbaum.
- MacWhinney, Brian and Catherine Snow
1990 The child language data exchange system: An update. *Journal of Child Language* 17, 457–472.

- Munakata, Yuko, James McClelland, Mark Johnson, and Robert Siegler
1997 Rethinking infant knowledge: Toward an adaptive process. *Psychological Review* 104, 686–713.
- Oshima-Takane, Yuriko
1988 Children learn from speech not addressed to them: The case of personal pronouns. *Journal of Child Language* 15, 95–108.
- Oshima-Takane, Yuriko, Yoshio Takane, and Thomas Shultz
1999 The learning of first and second person pronouns in English: Network models and analysis. *Journal of Child Language* 26, 545–575.
- Peters, Ann
1983 *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Pine, Julian, Elena Lieven, and Caroline Rowland
1998 Comparing different models of the development of the English verb category. *Linguistics* 36, 807–830.
- Pinker, Steven
1984 *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
1989 *Learnability and Cognition*. Cambridge, MA: MIT Press.
- Rice, Mabel, Kenneth Wexler, and Patricia Cleave
1995 Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research* 38, 850–863.
- Rice, Mabel, Kenneth Wexler, and Scott Hershberger
1998 Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 41, 1412–1431.
- Rice, Mabel, Kenneth Wexler, Janet Marquis, and Scott Hershberger
2000 Acquisition of irregular past tense by children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 43, 1126–1145.
- Richards, Brian
1994 Child-directed speech and influences on language acquisition: Methodology and interpretation. In Richards, Brian and Clare Gallaway (eds.), *Input and Interaction in Language Acquisition*. Cambridge: Cambridge University Press.
- Rowland, Caroline and Julian Pine
2000 Subject-auxiliary inversion errors and *wh*-question acquisition: “What children do know?”. *Journal of Child Language* 27, 157–181.
- Theakston, Anna, Elena Lieven, Julian Pine, and Caroline Rowland
2001 The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language* 28, 127–152.
2002 Going, going, gone: The acquisition of the verb “Go”. *Journal of Child Language* 29, 783–811.
- Theakston, Anna, Elena Lieven, and Michael Tomasello
2003 The role of the input in the acquisition of third person singular verbs in English. *Journal of Speech, Language, and Hearing Research* 46, 863–877.
- Tomasello, Michael
1992 *First Verbs: A Case Study of Early Grammatical Development*. Cambridge: Cambridge University Press.
2000 Do young children have adult syntactic competence? *Cognition* 74, 209–253.
- Tomasello, Michael, and Kirsten Abbot-Smith
2002 A tale of two theories: Response to Fisher. *Cognition* 83, 207–214.

Valian, Virginia

1991 Syntactic subjects in the early speech of American and Italian children. *Cognition* 40, 21–81.

Wexler, Kenneth

1994 Optional infinitive, head movement and the economy of derivations. In Hornstein, N. and D. Lightfoot (eds.), *Verb Movement*. New York: Cambridge University Press.

Offenlegung der Inhaber und Beteiligungsverhältnisse gem. § 7a Abs. 1 Ziff. 1, Abs. 2 Ziff. 3 des Berliner Pressegesetzes: Gisela Cram, Rentnerin, Berlin; Dr. Annette Lubasch, Ärztin, Berlin; Elsbeth Cram, Pensionärin, Rosengarten-Alvesen; Dr. Hans-Robert Cram, Verleger, Kleinmachnow; Margret Cram, Studienrätin i. R., Berlin; Verena Graß, Schülerin, Leimen; Brigitta Duvenbeck, Oberstudienrätin, Bad Homburg; Liselotte Schuchardt, Ärztin, Berlin; Dr. Georg-Martin Cram, Unternehmens-Systemberater, Stadtbergen; Jens Cram, Student, Stadtbergen; Renate Tran, Zürich; Gudula Gädeke M.A., Atemtherapeutin/Lehrerin, Tübingen; John-Walter Siebert, Pfarrer, Walheim; Dr. Christa Schütz, Ärztin, Mannheim; Dorothee Seils, Apothekerin, Stuttgart; Gabriele Seils, Journalistin, Berlin; Walter Cram, Architekt, Mexico DF (Mexiko); Ingrid Cram, Betriebsleiterin, Tuxpan/Michoacan (Mexiko); Sabina Cram, Mexico DF (Mexiko); Dr. Clara-Eugenie Seils, Oberstudienrätin i.R., Reppenstedt; Christoph Seils, Journalist, Berlin; Angelika Crisolli, kaufm. Angestellte, Hohenstein; Susanne Cram-Gomez, Mexico DF (Mexiko); Kurt Cram, Großhändler, Cancun (Mexiko); Silke Cram, Wissenschaftlerin, Mexico DF (Mexiko).