J. Child Lang. **30** (2003), 75–115. © 2003 Cambridge University Press
DOI: 10.1017/S0305000902005512 Printed in the United Kingdom

Lexically specific constructions in the acquisition of inflection in English*

STEPHEN WILSON
University of California, Los Angeles

(*Received* 26 *April* 2001. *Revised* 5 *June* 2002)

ABSTRACT

Children learning English often omit grammatical words and morphemes, but there is still much debate over exactly why and in what contexts they do so. This study investigates the acquisition of three elements which instantiate the grammatical category of 'inflection' – copula *be*, auxiliary *be* and 3sg present agreement – in longitudinal transcripts from five children, whose ages range from 1;6 to 3;5 in the corpora examined. The aim is to determine whether inflection emerges as a unitary category, as predicted by some recent generative accounts, or whether it develops in a more piecemeal fashion, consistent with constructivist accounts. It is found that for each child the relative pace of development of the three morphemes studied varies significantly, suggesting that these morphemes do not depend on a unitary underlying category. Furthermore, early on, *be* is often used primarily with particular closed-class subjects, suggesting that forms such as *he's* and *that's* are learned as lexically specific constructions. These findings are argued to support the idea that children learn 'inflection' (and by hypothesis, other functional categories) not by filling in pre-specified slots in an innate structure, but by learning some specific constructions involving particular lexical items, before going on to gradually abstract more general construction types.

[*] This research was supported by a University Fellowship from the UCLA Department of Linguistics. I would like to thank Ayşe Pınar Saygın, Bruce Hayes, Tim Arbisi-Kelm, Elena Lieven, Susan Curtiss, Nina Hyams, Carson Schütze, Ken Wexler, and two anonymous reviewers for very helpful comments and suggestions, and the audiences of talks at UCLA and UCSD where I presented some of this material. I thank also Roger Brown, Patrick Suppes and Jacqueline Sachs, for their generosity in submitting data to CHILDES which I have used, and Brian MacWhinney, for maintaining this invaluable resource. All responsibility for errors and shortcomings is of course entirely my own. Address for correspondence: Stephen Wilson, UCLA, 1320 Gonda Center, 695 Young Drive South, Los Angeles, CA 90095-1761, USA. e-mail: stephenw@ucla.edu

INTRODUCTION

It has been well known since the work of Brown and his colleagues (e.g. Cazden, 1968; Brown, 1973) that children learning English often omit grammatical words and morphemes in contexts where they are obligatory in the adult grammar. The following sentences exemplify missing copula *be*, missing auxiliary *be* and missing third person singular (3sg) present agreement, respectively:

(1) (*a*) I at the beach. (Nina, 2;5) (Suppes, 1974; MacWhinney, 2000)
(*b*) A lady dancing. (Nina, 1;11)
(*c*) The big doll need the bottle. (Nina, 2;3)

Brown (1973) showed that when children first begin to put words together, inflections, along with other grammatical elements, are missing often or even all of the time; this is what led to this early speech being labelled 'telegraphic'. The elements which are missing in (1) are used increasingly frequently through Brown's stages II–V, though they do not usually reach Brown's acquisition criterion of 90% correct in obligatory contexts until even later (Brown, 1973: 271). The aim of this study is to examine the patterns of development of the three grammatical elements which are missing in (1), especially *be* in (1*a*) and (1*b*), with respect to the predictions of several current theories.

Over the last decade there have been two dominant views within the generative tradition as to why inflections (and other functional elements) are often missing in the early speech of children learning English, and how they are acquired. One is the FULL COMPETENCE HYPOTHESIS (FCH) (Hyams, 1992*a*; Poeppel & Wexler, 1993) which claims that children have functional categories such as IP (inflectional phrase) right from the start, as a property of universal grammar. The absence of overt inflectional material in early speech is argued to be due to some specific property of child grammars, for instance, optional underspecification of functional heads such as Tense and Agreement (Schütze & Wexler, 1996; Schütze, 1997).

The second view which has been influential can be referred to as the STRUCTURE-BUILDING MODEL (SBM) (Aldridge, 1989; Radford, 1990, 1996; Vainikka, 1994). According to this view, functional categories are entirely absent in children's early grammars, and the sentences children produce are actually small clauses or VPs. Functional categories such as IP and CP (complementizer phrase) then develop at around the age of 2;0 or later, whereafter morphemes and lexical items which rely on these categories begin to be used.

Both of these approaches are strongly nativist, especially the full competence hypothesis, which presupposes detailed knowledge of clause structure from infancy, but also the structure-building model, where the development of functional categories is presumed to reflect either maturation (e.g. Radford, 1990) or the triggering of innate categories through lexical learning (e.g. Radford, 1996; see also Clahsen, Eisenbeiss & Penke, 1996 for a similar proposal).

In this paper, arguments are presented for a very different view: that children learn the functional category of inflection by learning some specific constructions which instantiate inflection, before gradually abstracting more general categories. Crucially, there is no IP primitive either present from early on, or triggered at some particular point, but rather a gradual accumulation of increasingly flexible constructions which contain inflections and other items held to depend upon IP, such as modals. The notion that children's knowledge of language proceeds from specific to general has been widely held within the traditional language acquisition literature (e.g. Braine, 1976; MacWhinney, 1982). Recent work has provided strong support for this position by presenting evidence suggesting initial lexical specificity in domains such as argument structure in English (Tomasello, 1992, 2000), subject–verb agreement in Brazilian Portuguese (Rubino & Pine, 1998), and question formation in English (Rowland & Pine, 2000); see also Lieven, Pine & Baldwin (1997) and Pine, Lieven & Rowland (1998) for further recent work along these lines.

The three positions outlined above make different predictions about patterns in the acquisition of inflection. We will start by discussing general predictions, before posing some specific questions and formulating more detailed predictions relating to these.

Both the FCH and the SBM assume that inflection is a psychologically real category for the child, so many versions of these theories predict that certain morphemes which instantiate inflection should emerge with some degree of parallelism, as they are all reflections of this underlying category. This prediction is made explicitly by Rice, Wexler & Hershberger (1998), working within a full-competence theory: 'the prediction is that growth curves for individual morphemes [which instantiate the abstract feature of Tense – *SW*] should be highly similar to each other and to a composite measure' (p. 1417). It is important to note that most theories include various factors which could produce different rates of acquisition among different morphemes.[1] But there are cases in which current theories explicitly claim that realization of more than one inflectional morpheme depends upon exactly the same underlying element(s), and in these cases, parallel emergence is predicted. These cases will be discussed in more detail shortly.

A second key prediction of both the FCH and the SBM is that once children have begun to use a particular functional word or morpheme, they should use it productively. This follows from the fact that appearance of inflections is taken to reflect knowledge of IP, which should therefore allow the child to use the inflection in all appropriate lexical contexts. The SBM is actually somewhat

[1] For instance, the separation of Agr and Tns in Schütze (1997) predicts that there should be a stage of development during which the past tense suffix is used more frequently in obligatory contexts than the present tense suffix, since it requires only Tns to be specified, whereas the present tense suffix relies on both Tns and Agr.

more complicated than this: Radford (1990) argues that early uses of inflection (before age 2;0 or so) are in fact non-productive, which is necessary for his claim that these children have no functional categories. However, when IP is acquired, inflections should become productive. In short, many recent generative models predict PARALLELISM and PRODUCTIVITY in the emergence of inflectional material.

In contrast, this study argues for the constructivist view that children learn inflection by learning a range of constructions which happen to contain inflectional material. This approach therefore predicts pervasive asymmetries in the developmental schedules of the various constructions instantiating inflection. Variation would be expected both across morphemes (for example, in the relative rates of acquisition of copula *be* and 3sg present agreement) and across lexical contexts (for example, whether *be* appears or not might be expected to depend on what particular lexical item is in the subject position).

We are now in a position to pose two specific empirical questions which are addressed in this study:

(I) What are the relative rates of acquisition of copula *be*, auxiliary *be* and 3sg present agreement *-s*?
(II) What, if any, is the effect of the subject (whether it is open- or closed-class, and which closed-class pronoun or demonstrative it is) on the realization of copula and auxiliary *be*?

The first question is an obvious one to ask in order to test whether morphemes which are assumed to depend upon IP emerge in parallel, as predicted by some versions of the FCH and the SBM, or whether there are asymmetries, as would follow from the constructivist account. Rice *et al.* (1998) test their prediction, quoted above, that these morphemes emerge in parallel. They do find a fair degree of parallelism with several morphemes which instantiate tense, especially in comparison to the plural morpheme which is acquired much earlier. However, they report only results summed across children, so it is not possible to see whether the morphemes develop in parallel for each individual child. Also, they do not report any statistical tests which would reveal whether or not the summed growth curves for morphemes they do provide are significantly different from one another.

The second question is designed to test lexical specificity in the early use of inflectional material, which is predicted by the constructivist account but not by the FCH or the SBM. Consider the hypothetical examples in (2)–(4):

(2) (*a*) He big.
(*b*) He's big.

(3) (*a*) That big.
(*b*) That's big.

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

(4) (*a*) The truck big.
(*b*) The truck's big.

The question is whether there is any difference in the relative frequency of the (*a*) and (*b*) options in cases like (2) and (3), where the subjects are closed-class, compared to (4), where the subject is open-class. It can also be asked whether there are differences among particular closed-class subjects. For example, does the relative frequency of the (*a*) and (*b*) options differ according to whether the subject is *he*, as in (2), vs. *that*, as in (3)?

The motivation for this particular question is that it allows a particularly clear comparison between the theories under consideration. In the constructivist model, forms such as *he's* and *that's* would be expected to often be learned as 'chunks', or lexically specific constructions. The FCH, on the other hand, takes such forms as evidence for the child having an IP category, and therefore predicts that *be* should also occur productively in similar environments, such as with different pronominal subjects or with lexical NP subjects. The SBM makes the same prediction, once children have passed out of the initial stage of no functional categories. The following sections discuss in more detail the predictions of the various theories with respect to these specific questions.

Full competence theories

All versions of the FCH assume that young children have innate knowledge of the full clausal architecture which is assumed to be part of universal grammar. In most theories, it is also assumed that at the stage of development with which we are concerned, lexical learning of the relevant inflectional morphemes has already taken place. This claim is based on the fact that when children do use inflectional material, they almost always do so correctly (Brown, 1973; see Hyams (1999) for references documenting this finding in many languages). Hyams (1999) claims that '[g]iven that children use agreeing forms of the verb with a high degree of accuracy, it cannot be the case that root infinitives [i.e. missing inflections, for present purposes – *SW*] arise from a lack of knowledge of the specifier–head agreement requirement OR OF THE SPECIFIC FORMS THEMSELVES' (p. 395, emphasis added). A similar argument is made by Wexler (1998: 41–43).

Given these assumptions, it is necessary for proponents of the FCH to explain why children produce so many non-adult utterances. Most accounts address this by assuming that there are one or more grammatical or pragmatic factors which systematically cause or permit the omission of inflectional material. Exactly what these factors are vary from theory to theory; here we discuss three theories.

The Agr/Tns Omission Model (Schütze & Wexler, 1996; Schütze, 1997) claims that inflectional material is missing from children's early utterances

because the functional heads Tns (tense) and Agr (agreement) (into which IP is divided) can be optionally, independently, underspecified in any given utterance. The three morphemes of interest in this study – (finite) copula *be*, (finite) auxiliary *be* and 3sg present agreement – all mark both tense and agreement, therefore according to the Agr/Tns Omission Model, each of them will appear only when Tns and Agr are specified. If either Tns or Agr are underspecified in a particular utterance, *be* or 3sg agreement will be omitted. It therefore follows under this model that copula *be*, auxiliary *be* and 3sg present agreement should be acquired at the exact same rate, since all three rely on exactly the same abstract elements: Tns and Agr. Throughout development, each morpheme would be expected to be used in the same proportion of obligatory contexts. As noted above, this prediction is made explicitly by Rice *et al*. (1998) working within a closely related theory. There is also no obvious mechanism in this model by which lexical context could play a role in determining appearance or non-appearance of inflections, since the factors taken to be responsible for missing inflections are abstract grammatical properties.

Hyams (1999) proposes that inflections are omitted because children lack a certain pragmatic principle which would force them to supply finite forms. It is claimed that '[f]initeness becomes obligatory when the relevant pragmatic principle develops' (p. 403). There is no claim in this model that morphemes should develop in parallel, because in other work, Hyams has suggested that factors such as modality can play a role in determining omission (e.g. Hoekstra & Hyams, 1998), which would imply that different developmental patterns across constructions could be expected to the extent that these constructions differ in their modal properties. However, the model implicitly rules out effects of lexical context, since it is claimed that the specific forms of relevant morphemes are known, and omission depends upon the lack of a pragmatic principle.

A third theory (Becker, 2000) is especially relevant in the context of the present study since it focuses on the development of copula and auxiliary *be*. Becker claims that in copular constructions, predicate type (nominal, adjectival or locative) is the major determinant of whether or not *be* will be omitted. Becker observes that most nominals and some adjectives are individual-level predicates (i.e. they denote permanent properties), whereas most locatives and some adjectives are stage-level predicates (denoting temporary properties). Her proposal is that stage-level predicates contain an additional aspectual projection which, at one developmental stage, results in these clauses being realized as non-finite. Therefore she predicts that during this stage, the copula will always be present with nominal predicates, will never be present with locative predicates, and will sometimes be present with adjectival predicates. The data to be presented below suggest there may be some asymmetries among these categories of predicate, but the differences do not appear to be as sharp as Becker claims, nor do they necessarily go in the directions she

predicts. The present progressive overtly marks aspect and so must also contain an aspectual projection just as stage-level predicates do, which should result in the auxiliary *be* being systematically omitted. Becker observes that this is not the case, and accepts that she has no explanation for why this prediction is not borne out (p. 146). She also leaves unresolved the question of aspectual properties of main verbs, so it is not clear what predictions her account makes for 3sg present agreement (pp. 140–44). In sum, Becker's account predicts that copulas will be overt more frequently than auxiliaries, and leaves the door open for other differences between types of inflection since they will differ in their aspectual properties. However, in common with the other theories, her account rules out lexical context as a factor. In fact, she presents data which she claims rule out subject type as an important determinant of omission (p. 150). Later, the data which underlie this claim will be discussed.

To summarize the predictions of these full competence theories, some accounts predict parallel emergence of the morphemes examined in this study (e.g. Schütze, 1997; Rice *et al.*, 1998), whereas others propose omission mechanisms which could have different effects on different morphemes (e.g. Hyams, 1999; Becker, 2000). None of these accounts appear to offer any mechanisms which would permit effects of lexical context on the realization of inflections.

It should be noted that a rather different version of the FCH in which something akin to lexical learning plays a major role has been proposed by Phillips (1995). This work, which is in some respects closer to the constructivist account offered in this study, will be returned to briefly in the discussion.

The structure-building model

We turn now to the structure-building model. Radford (1990, 1996) argues that children learning English go through a stage where their grammars contain only lexical categories and no functional categories, roughly between the ages of 1;8 and 2;0. (See Aldridge (1989) and Vainikka (1994) for work along the same lines.) Radford (1990) argues that children's clauses at this time are verb phrases (with subjects in Spec of VP), and that there is no productive use of determiners, inflections, complementizers, or anything that would rely on functional projections such as IP or CP. He argues that when these elements do appear at this early stage, either they are very limited in productivity, or they should be given a different status in the child grammar than they will eventually have in the adult grammar.

The most important question for present purposes is how, under this model, these early clauses subsequently develop into fuller functional structures. Radford (1990) suggests that both IP and CP 'mature' simultaneously at

about 2;0, and claims that after this age children very quickly acquire the full range of functional structures which they have been missing. In Radford (1996), he outlines an alternative, which is that IP appears first (at about 2;0) and CP somewhat later (at about 2;6). It is less clear in this later paper whether Radford thinks that maturation or lexical learning is involved, but his claim that languages with a richer verbal morphosyntax might 'force the child learning such a language to project an IP constituent at the outset' (1996: 65) would seem to be inconsistent with a maturational explanation.

One of the most serious issues for an account such as this is that children do not move through clearly delineated stages where they produce only VPs, then only IPs, then only CPs. This is acknowledged by Radford (1996), who observes that 'when they first acquire a given type of functional extended projection, children only optionally project the relevant functional architecture' (1996: 67). For instance, children who are claimed to have acquired IP and even CP continue to produce significant numbers of VP clauses. Radford offers no substantive explanation for why this should be the case, except to state that children 'optionally TRUNCATE structures' (1996: 75, emphasis in original).

With respect to questions (I) and (II) posed above, Radford's (1990) model predicts rapid and across-the-board acquisition of elements dependent on IP. In contrast to this, Radford (1996) backs away from the claim of rapid development but offers only an unspecified optional truncation process to explain missing inflections. The model in Radford (1996) thus does not offer any mechanism by which to account for asymmetries in development across constructions (question I) or in particular lexical contexts (question II).

Note that this study does not address Radford's claim that very young children have no knowledge of functional categories. The children whose transcripts are examined have either passed out of Radford's early stage, or pass out of it early in the time period examined (*cf.* Vainikka, 1994), so this is not an issue which can be addressed here. The point of contention, rather, is whether there is any evidence for the triggering or maturation of an IP category, or whether inflection is something which is learned piece by piece.

A constructivist account

In this section we outline a constructivist account, whereby learning the grammatical category 'inflection' relies upon first learning specific constructions. Following recent work in CONSTRUCTION GRAMMAR, we will assume that children's (and adults') grammatical knowledge takes the form of CONSTRUCTIONS (e.g. the simple active, the ditransitive, the resultative) (Fillmore, Kay & O'Connor, 1988; Goldberg, 1995; Tomasello, 2000). Two features of construction grammar are especially important in this study. One is the idea that a more satisfactory account of linguistic competence can be obtained if

constructions are taken as genuine psychological entities. This contrasts with the view of the principles and parameters framework where traditional constructions are seen as epiphenomenal, arising from the interaction of general principles. The second important claim is that there is no sharp dividing line between grammar and the lexicon. Psycholinguistic and neurolinguistic evidence for this position is reviewed by Bates & Goodman (1999). In the present study it is argued that because of this continuity, grammatical constructions show some of the same properties as lexical items in terms of how they are acquired: specifically, although the time at which they are acquired is influenced by factors such as complexity and frequency, there is also a very significant amount of individual variation across children.

In construction-based approaches to acquisition, early constructions revolve around particular lexical items (Braine, 1976; Tomasello, 1992), becoming increasingly abstract as the grammar develops. Peters (1983; see also Peters, 1997 and the references therein) develops a detailed and empirically well-motivated account of how grammatical development might proceed in this kind of model. She discusses the extraction of units of various sizes from the speech stream, the segmentation of extracted units into smaller ones, the formation of frames which contain a variable part, and the progressive generalization of these into yet more productive syntactic patterns. Similar proposals include Maratsos & Chalkley (1980), MacWhinney (1982), Tomasello (1992, 2000) and Tomasello & Brooks (1998). Further evidence for lexical specificity in early grammar, and/or arguments along similar lines, are presented by Kuczaj & Maratsos (1983), Lieven *et al*. (1997), Pine *et al*. (1998), Rubino & Pine (1998) and Rowland & Pine (2000).

We will now discuss some of the specific constructions which are proposed here to underlie children's early sentences involving both present and absent inflections. In order to account for the omission of functional morphemes, it crucially needs to be assumed that children initially derive constructions which are 'pared down' versions of the adult models to which they are exposed. Presumably there are many factors leading to function words being the ones to go, including relative lack of informativeness (Greenfield & Smith, 1976), greater semantic and grammatical complexity (Brown, 1973) and, probably less importantly, lack of acoustic salience (Gleitman & Wanner, 1982). Brown (1973) discusses factors such as these in an attempt to identify determinants for his findings on the order of acquisition of grammatical morphemes, concluding that semantic and grammatical complexity are probably the most important. These issues will not be addressed here, but it seems reasonable to assume that the preference to omit function words rather than content words can be derived from more basic factors such as these.

To take present progressives for an example, it is proposed that after hearing large numbers of sentences such as *he's running*, *you're climbing*, *the*

cat's sitting, *I'm eating* and so on, the child will abstract a construction of the form:

(5) NP_{subj} V-ing.

This initial construction contains just the most essential items; the auxiliary is omitted. This construction would be relied upon in producing sentences like *he running*, *you climbing*, *the cat sitting* and *I eating*. It is worth considering another logical possibility for a pared down present progressive construction: NP_{subj}'s V, with *-ing* omitted; the child would then produce sentences such as *he's run*, *the cat's sit*, and so on, in present progressive contexts. Empirically this cannot be correct, as such sentences are rarely found. There are probably several factors which make *-ing* more salient and easier to incorporate into a construction than copula/auxiliary *be*: it is always a syllable, its vowel is not reduced, its form is constant across subjects, and it has identifiable semantic content (progressive aspect), which is not the case for *be*, occurring as it does in numerous varied contexts.

At the same time as the child is using the NP_{subj} V-*ing* construction in (5), he/she will also be learning to produce *be* in similar sentences. Closed-class forms such as *he*, *she* and *I* occur very frequently as subjects, which means that it is feasible for the child to abstract lexically specific constructions for closed-class subjects which include the appropriate allomorph of *be*:

(6) (*a*) he's V-ing
(*b*) you're V-ing
(*c*) I'm V-ing

These constructions are proposed to be similar to lexical items in many ways. Specifically, it is expected that children will vary in terms of which particular subjects they include in constructions like those in (6). These constructions co-exist in the grammar with the more general but more basic NP_{subj} V-*ing* construction in (5). Again it is worth considering what might motivate the particular segmentations proposed here. Forms such as *he's*, *you're* and *I'm* constitute prosodic words. This may lead to them being more readily extractable units than other recurring sequences such as *is* V-*ing*, which does not constitute a single prosodic word.

Any particular open-class subject, such as *the pony*, presumably occurs much less frequently than any closed-class subject, so it is proposed that it is much less feasible for the child to abstract constructions such as *the pony's* V-*ing*. However it is plausible that some high-frequency lexical subjects such as *Mommy* and *Daddy* might also be learned as units along with *be*. Therefore, to be precise, the claim is not that there is an inherent difference between open- and closed-class subjects in terms of whether they can be chunked with *be*, but rather, chunking should occur much more often with closed-class subjects than it does with open-class subjects.

In proposing that the child's constructions may include elements such as *he's* and *I'm*, which are composed of two morphemes, this account follows Peters (1983) in suggesting that the basic units for the child often do not coincide with the morphemes which are generally assumed to be the basic units in the lexicon. The position that items like *he's* and *I'm* may be unanalysed in child grammar has been held by many researchers. However, an important point needs to made. Although we will argue that *he's* and *I'm* are often unsegmented in child grammars, this does not imply that they are simply equivalent to *he* and *I*, as some researchers have seemed to suggest (e.g. Pinker, 1996: 261). Empirically it is clear that they are not, because it is very rare that children say things like *I'm want it*, which would be expected if they did not distinguish between *I'm* and *I*. In terms of the present account, *I* and *I'm* are claimed to be represented very differently in the child's grammar: the unit *I'm* exists only as part of the construction in (6*c*), and other construction(s) for copula sentences. It has no independent existence as a lexical item which would allow it to be used to construct a sentence like *I'm want it*.

So far, we have proposed that the child's grammatical knowledge comprises general but pared down constructions such as (5), and specific constructions containing inflections such as those in (6). In order to produce the present progressive productively however, the child must derive constructions which are both general in terms of the subject slot, and which also contain *be* overtly:

(7) (*a*) $NP_{subj\text{-}3sg}$'s V-ing
(*b*) $NP_{subj\text{-}3pl}$'re V-ing

Crucially, these constructions are predicted to develop later than those in (5) and (6). This follows from the assumption that complex constructions are derived by abstraction across simpler constructions (Peters, 1983). So $NP_{subj\text{-}3sg}$'s V-*ing* could be derived by abstracting away from *he's* V-*ing*, *she's* V-*ing*, *Mommy's* V-*ing*, and so on. That is, the constructions in (7) are derived from those in (6) and so should develop later. Note that the subjects in (7) are restricted in terms of person and number. This encodes the theoretical assumption that children are conservative about the generalizations they draw, and captures the empirical fact that children rarely say things like *I's going*, which could be produced if the subject in (7*a*) were not appropriately restricted.

So at the stage when inflections are being used sometimes but not always, the child has several overlapping constructions, varying in specificity, by which to express present progressive meanings. It will be assumed that competence with copulas develops similarly, though here there are also demonstratives, as well as pronouns, which can be learned as chunks: *that's* XP_{pred}, *this is* XP_{pred}. With regard to 3sg present forms, this approach implies that early forms would rely on verb-specific constructions such as $NP_{subj\text{-}3sg}$ *likes* NP_{obj}, but this is not

tested in the present study, because only one of the five children examined produced a large number of 3sg present agreement morphemes.

In sum, the constructivist account predicts that there will be significant individual differences in the relative rates at which copula *be*, auxiliary *be* and 3sg present agreement are produced, because these are all different constructions, and are hence subject to the randomness inherent in the learning of distinct items. Copulas can be predicted to develop somewhat earlier than auxiliaries simply because the auxiliary construction is more complex by virtue of containing one more element: the *-ing* suffix. Valian (1992) makes a very similar argument within a nativist approach. Aside from that, no further predictions are made about exactly what relative developmental rates would be expected for the three morphemes. The claim is only that they would be expected to often differ significantly, and inconsistently, since it is assumed that the knowledge underlying each construction is distinct.

The constructivist account predicts that on the whole, copula and auxiliary *be* should occur more frequently with closed-class (or highly frequent) subjects with which *be* can be learned as a chunk. If subject-*be* combinations really are learned as chunks, then random differences in the production of *be* with different subjects are also expected, just as would presumably be found with the acquisition of any lexical items. We will now set out to test the predictions of the various models with respect to questions (I) and (II) posed above.

METHOD

Transcripts taken from the CHILDES database (MacWhinney, 2000) of five children learning English were studied: Adam, Eve and Sarah (Brown, 1973), Nina (Suppes, 1974) and Naomi (Sachs, 1983). These transcripts were selected because of the large amounts of longitudinal data available for these children. Table 1 provides basic information about the data: which files were coded, and the age range and mean length of utterance in words (MLU_w) for each child. The transcripts for each child were split into four time periods prior to analysis for the purpose of carrying out analyses with time as a factor. The divisions were made so as to roughly match the amount of time and amount of transcript material for each period, without regard to how the child's linguistic development was proceeding. These groupings are also shown in Table 1.

Children's utterances were coded according to the following criteria: only declarative sentences were coded. Both positive and negative sentences were included, as were statements with tag questions, though the tag itself was not counted. Interrogatives of all types (including forms with no overt movement, e.g. *this is a car?*) and imperatives were excluded, simply so as to reduce the number of factors which need to be considered. Exact or reduced imitations of recent adult utterances, and self-imitations, were excluded, as were songs, stories, and so on. Utterances containing unclear material were included if the

TABLE 1. *Longitudinal data used for the study*

Child	Period 1	Period 2	Period 3	Period 4	Total (utterances)
Adam					
Files	1–8	9–16	17–23	24–30	1–30 (26 077)
Ages	2;3–2;6	2;6–2;10	2;10–3;1	3;2–3;5	2;3–3;5
MLUw	2·26	2·51	3·30	3·74	2·94
Eve					
Files	1–5	6–10	11–15	16–20	1–20 (11 208)
Ages	1;6–1;8	1;9–1;10	1;11–2;1	2;1–2;3	1;6–2;3
MLUw	1·74	2·57	3·13	3·26	2·73
Sarah					
Files	25–33	34–42	43–51	52–60	25–60 (8491)
Ages	2;8–2;10	2;10–3;0	3;1–3;3	3;3–3;5	2;8–3;5
MLUw	2·11	2·27	2·65	2·63	2·41
Nina					
Files	1–7	9–15	16–22	23–31	1–31 (42 074)
Ages	1;11–2;0	2;1–2;2	2;3–2;4	2;4–2;5	1;11–2;5
MLUw	2·03	2·44	3·11	2·92	2·60
Naomi					
Files	3–19	20–35	36–51	52–68	3–68 (10 056)
Ages	1;8–1;10	1;10–2;0	2;0–2;2	2;3–2;7	1;8–2;7
MLUw	1·58	2·19	2·12	3·04	2·18

The figures in parentheses following the total file ranges show the total number of child utterances in these files.

region of interest was clear. If a child produced exactly the same utterance more than five times in the corpus, it was only counted the first five times. This resulted in the exclusion of just a few presumably wholly formulaic phrases such as *that's right* and *you're welcome*, which might otherwise have skewed the results.

Three basic environments where the adult grammar requires overt inflectional material were coded: present tense copular constructions; present progressives; and sentences requiring 3sg present agreement.

Copular constructions were coded for subject type, presence vs. absence of the copula, and type of predicate. Predicate types were divided into nominal (e.g. *he's a dog*), adjectival (*she's happy*), locative (*it's in the kitchen*) and 'other' (including possessives and participles).

Present progressives were defined as clauses with *-ing* on the verb which seemed according to context to have present reference. These were coded for subject type and for presence vs. absence of the auxiliary. Sentences such as *he's* V, which could plausibly be progressive constructions with missing *-ing*, were not counted as present progressives. These proved to be quite rare and their counts are not reported in this paper.

For both copular and progressive constructions, clauses with missing subjects were excluded, as these would be extremely difficult to judge. Clauses

with non-nominative subjects (e.g. *me a boy*) were also excluded; only one child produced more than a handful of these (see note 3). Past tense copulas and progressives were not counted; these are also infrequent at this stage of development.

Both contracted and uncontracted copulas and auxiliaries were counted. In those environments where contraction of *be* can be indicated in the orthography, the children were observed to nearly always use the contracted form. However, counts for contracted vs. uncontracted forms are not reported. There are some syntactic contexts in which *be* can never be contracted, usually involving VP ellipsis (e.g. *who's a girl? I am!*). These types of utterances, which are relatively infrequent, were excluded, since according to the present account they might well rely upon separate constructions.

3sg present agreement was coded as present when it appeared in an appropriate context, and absent when the context made it clear that it would have been called for in the adult grammar. In practice, as has often been noted, such contexts can be somewhat difficult to determine. When a child says *he go*, it is difficult to know whether they 'intended' *he goes*, *he's going*, *he went* or perhaps even *he could have been about to go*. However, the surrounding material usually makes it possible to determine whether 3sg agreement would have been required. Because these coding decisions were the most difficult, 15 files – three from each child – were coded by a second coder for presence vs. absence of 3sg agreement. The rate of agreement between the author and the second coder was over 95% (of all utterances containing verbs which did not have overt first or second person subjects, and thus were potentially relevant).

Children do produce occasionally agreement errors with copulas, auxiliaries and 3sg agreement. These errors were not counted or analysed because there are very few of them, but future work with denser transcripts will probably be able to derive crucial insights from such errors.

Several methodological limitations of this study should be noted. Firstly, because the transcripts used were fairly broadly transcribed, no attempt was made to assess whether the children's developing phonological systems contributed significantly to the patterns observed. Secondly, prosodic structure was also not taken into account, though several studies have provided evidence that prosodic structure is a factor in the omission of functional words and morphemes (e.g. Gerken, 1996). And thirdly, the reliance on transcripts made it impossible to investigate whether the children in this study used 'filler syllables' on their way to developing inflections (see Peters, 2001, for review).

RESULTS AND DISCUSSION

First, results are presented showing the overall course of development of the inflections of interest for each of the five children. Then comparisons are made to determine what effect the subject has on whether copulas and auxiliaries

TABLE 2. *Overall copula, present progressive and main verb agreement marking*

	Copula *be*			Auxiliary *be*			3sg agreement			Significance		
Child	Abs	Pres	%	Abs	Pres	%	Abs	Pres	%	Cop/Aux	Cop/3sg	Aux/3sg
Adam	656	518	44	871	146	14	220	134	38	>***	>*	<***
Eve	530	120	18	329	20	6	80	49	38	>***	<***	<***
Naomi	68	281	81	115	155	57	10	36	78	>***	>	<**
Nina	275	1052	79	264	187	41	245	38	13	>***	>***	>***
Sarah	212	254	55	87	32	27	66	37	36	>***	>***	<

Abs, absent; Pres, present; %, percent correct in obligatory contexts; Cop, copula *be*; Aux, auxiliary *be*; 3sg, 3sg present agreement. In the significance columns, the > or < shows which of the two categories was more often used correctly in obligatory contexts. The asterisks show which comparisons reached significance at * $p \leqslant 0{\cdot}05$, ** $p \leqslant 0{\cdot}01$ or *** $p \leqslant 0{\cdot}001$.

are present or absent. The next section looks at variation in the rate of overt *be* among individual closed-class subjects, and finally the possible confound of predicate type is investigated.

Overall course of development

The children's use of copula *be*, auxiliary *be* and 3sg present agreement in obligatory contexts is summarized in Table 2, and the development of these morphemes over time is shown in Figure 1. It can be observed that each child tends to use the inflections more frequently in obligatory contexts as he or she develops. But the rates of development are clearly different for the three morphemes. Significance was measured using Fisher's Exact Test, two-tailed. All five children used copula *be* more frequently than auxiliary *be* ($p \leqslant 0{\cdot}001$ for each child). This may be due to the copula construction being simpler in that it involves one fewer morpheme. The faster development of copula *be* relative to auxiliary *be* has been reported in several studies (e.g. Brown, 1973; Ingram, 1974; Valian, 1992).

However, the frequency of 3sg present agreement relative to copula *be* varies dramatically from child to child. For three of the children, the copula was significantly more frequent than 3sg agreement (Nina, Sarah, $p \leqslant 0{\cdot}001$; Adam, $p \leqslant 0{\cdot}05$), but for one, Eve, the opposite pattern was found, with 3sg agreement significantly more frequent than the copula ($p \leqslant 0{\cdot}001$). For Naomi, there was no significant difference. The comparison between Nina and Eve is particularly striking: Nina used the copula 79% of the time and main verb agreement just 13%, whereas Eve used the copula just 18% of the time but main verb agreement more than twice as often: 38% of the time.

This 'double dissociation' appears to be problematic for some current generative accounts. The fact that Nina correctly used the copula 79% of the time indicates that whatever putative factor licenses missing inflections could

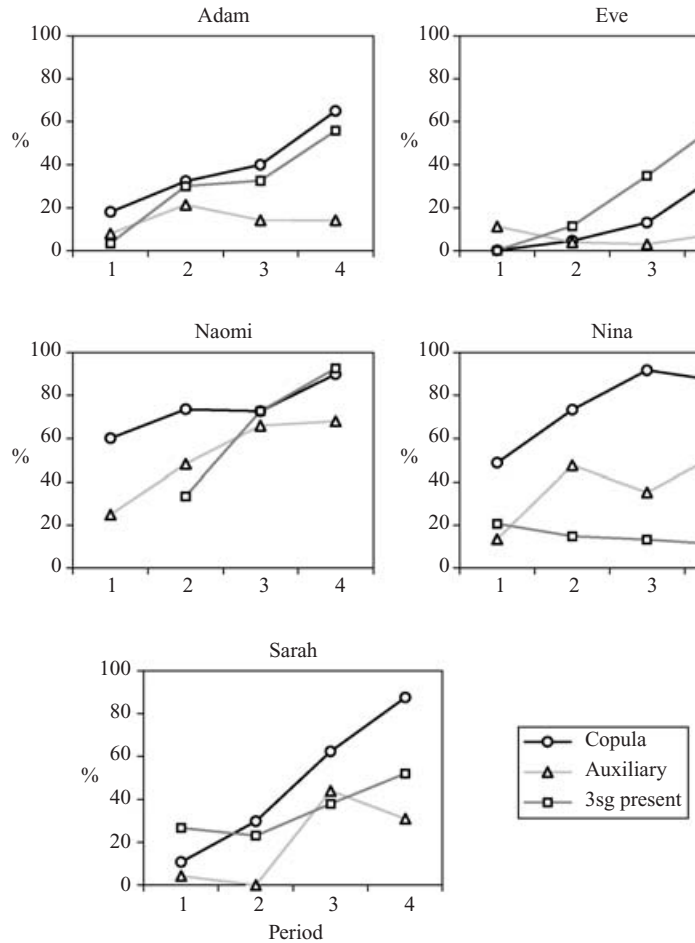

Fig. 1. Copula *be*, auxiliary *be* and 3sg present agreement marking by time. The *x*-axis shows time period and the *y*-axis shows percent correct in obligatory contexts. In these and all other figures in this paper, data points are shown only when there were at least four obligatory contexts contributing to the point.

only have been operating in a minority of utterances. It then becomes difficult to explain why she used main verb agreement only 13% of the time, if she is assumed to have full competence or, under a structure-building account, to have acquired IP (an assumption which would be forced by her reliable copula use). Furthermore, the data from Eve suggest that it would not be possible to propose a grammatical explanation for why 3sg agreement occurs so much less frequently than the copula, because Eve shows exactly the opposite pattern.

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

The only solution is to claim that these two constructions depend on distinct factors. This, of course, is exactly the constructivist position: it is claimed that at this early stage of grammatical development, constructions are to a large extent independent, and a more general category of inflection will only emerge at some later stage.

Turning to the comparison between auxiliary *be* and 3sg agreement, similar differences between children can be seen. Three of the children used 3sg agreement more frequently than auxiliary *be* (Adam, Eve, $p \leqslant 0{\cdot}001$; Naomi, $p \leqslant 0{\cdot}01$), but Nina shows the opposite pattern, with 41% correct auxiliary use but only 13% main verb agreement ($p \leqslant 0{\cdot}001$). For Sarah, no significant difference was found.

In sum, these data do not offer any support for the idea that children's production of the three morphemes under investigation relies upon a single underlying category (e.g. Rice *et al.*, 1998). Rather, the dissociations observed support the constructivist position that at this early stage of grammatical development there are at least three distinct constructions. In fact, it will be argued in the following sections that there are more than three, because production of a morpheme such as the copula relies on more than one construction.

Open- and closed-class subjects and the realization of be

The constructivist account predicts that copula and auxiliary *be* should tend to be more frequent after closed-class subjects, since in those cases the copula/auxiliary can be learned as a unit along with the subject (e.g. *he's*, *that's*), whereas for lexical NP subjects this is presumably less feasible.

To test this prediction, the proportion of the time copula and auxiliary *be* were present in obligatory contexts was compared for closed-class vs. lexical NP subjects. Since open-class subjects are virtually always third person, only third person closed-class subjects were included in this section, in order to provide an appropriate comparison.[2]

The overall counts for copulas are shown in Table 3, with development over time shown in Figure 2. The prediction of the constructivist account is borne out: four of the five children produced the copula significantly more frequently with closed-class subjects than with lexical NP subjects (Adam, Naomi, Nina, Sarah, $p \leqslant 0{\cdot}001$; Fisher's Exact Test, one-tailed). The overall differences in rate between the two conditions range from 34% to 47% for

[2] It has been claimed by Hyams (1999) that NPs with missing determiners are non-finite and do not license finite verb inflections. If this contingency turns out to hold up empirically after closer investigation, then it would contribute to a lower rate of overt agreement with lexical NP subjects, since some of them lack determiners at this stage of development. To guard against this possible confound, the analyses in this section were also carried out excluding all non-finite NPs. Very similar results were obtained to those presented.

TABLE 3. *Copula marking with closed-class and lexical NP subjects*

	Closed-class subjects (3sg/3pl)			Lexical NP subjects			
Child	Absent	Present	%	Absent	Present	%	Significance
Adam	428	469	52	103	23	18	>***
Eve	428	102	19	74	12	14	>
Naomi	29	238	89	31	22	42	>***
Nina	126	967	88	102	73	42	>***
Sarah	164	225	58	21	5	19	>***

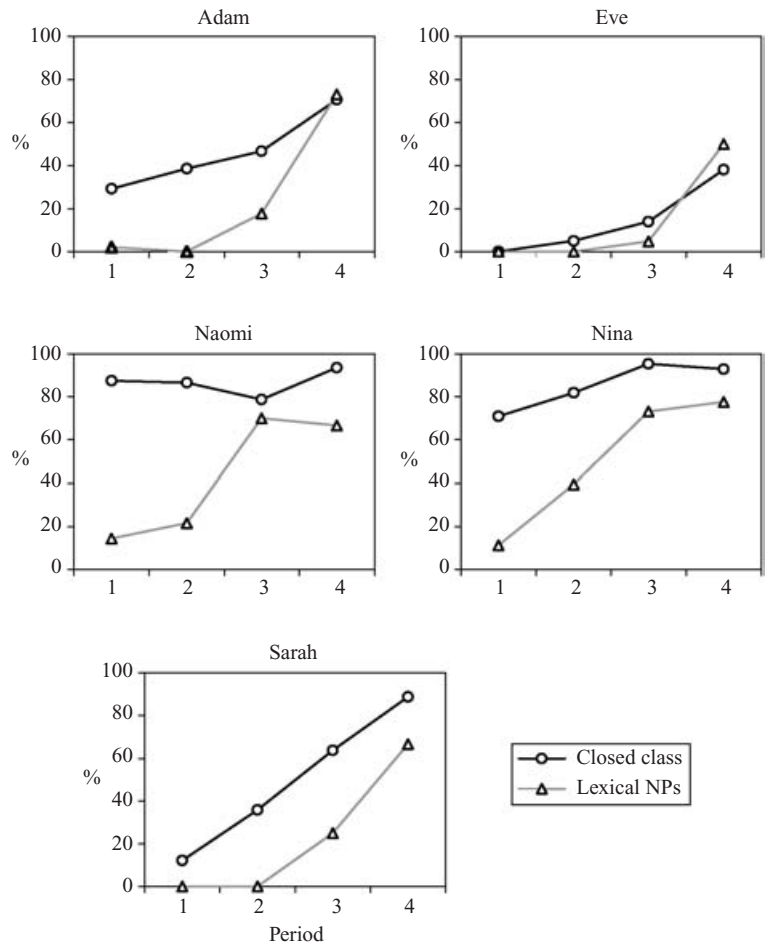

Fig. 2. Copula marking with third person closed-class and lexical NP subjects by time.

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

these four children. The fifth child, Eve, produced the copula slightly more frequently with closed-class subjects, but this difference was not significant.

Especially notable is the extremely low frequency of the copula with lexical NP subjects in the first half of the data for three of the children. Two of the children did not use a single copula with lexical NP subjects during that time: Eve had 43 obligatory contexts and Sarah had 16. Adam used just 1 copula in 82 obligatory contexts. In the same period, each of these children produced copulas with closed-class third person subjects (Adam, 89/252; Eve, 7/161; Sarah, 28/122). These data suggest that during this period, Adam, Eve and Sarah had no productive means of producing the copula, but were only capable of producing it (some proportion of the time) when it had been learned as a unit along with a closed-class subject. This pattern in Adam's and Eve's data was in fact observed by Brown (1973: 383) who arrived at a similar interpretation. As can be seen in Figure 2, in the first half of the data, Nina and Naomi also produced the copula much more frequently when it could be chunked, but they seem in addition to have been beginning to use it productively with all types of subjects. In terms of the constructions proposed earlier to underlie children's production of inflection, these data suggest that at this stage Adam, Eve and Sarah were operating with only constructions like those in (5) and (6), whereas Nina and Naomi were also using these types of constructions, but had also begun to abstract more complex constructions like those in (7).

Turning to auxiliary *be*, a similar pattern can be observed, with *be* more frequent after closed-class subjects, though the results are not quite as strong. The overall counts are given in Table 4 and the development of auxiliary *be* marking over time is shown in Figure 3. Three of the five children used auxiliary *be* significantly more frequently with closed-class subjects (Adam, Nina, Naomi, $p \leqslant 0{\cdot}001$; Fisher's exact test, one-tailed). For these three children, the percentage difference between the two conditions varies from 29% to 50%. For Eve and Sarah there are nonsignificant tendencies in the same direction. Note that the apparent drop over time in Adam's rate of supplying *be* with closed-class subjects probably just reflects sampling error; there were only 5 obligatory contexts in the second quarter of his files.

The same three children who produced only 1 copula between them with a lexical NP subject in the first half of the data were similarly limited in their abilities with auxiliaries with lexical NP subjects over that period. Adam and Sarah produced none, in 41 and 5 obligatory contexts respectively. Eve produced 2 in 50 obligatory contexts, but these were probably formulaic, since it was the exact same utterance (*fish are swimming*) in the two cases, which did not occur in the same file.

In sum, then, four of the five children used copula *be* significantly more frequently with closed-class subjects, and three of the five used auxiliary *be* significantly more frequently with closed-class subjects, consistent with the constructivist account in which *be* can be produced in these cases, but not in

TABLE 4. *Auxiliary marking with closed-class and lexical NP subjects*

	Closed-class subjects (3sg/3pl)			Lexical NP subjects			
Child	Absent	Present	%	Absent	Present	%	Significance
Adam	139	97	41	102	14	12	>***
Eve	48	5	9	94	5	5	>
Naomi	5	24	83	55	38	41	>***
Nina	39	134	77	111	42	27	>***
Sarah	30	13	30	10	3	23	>

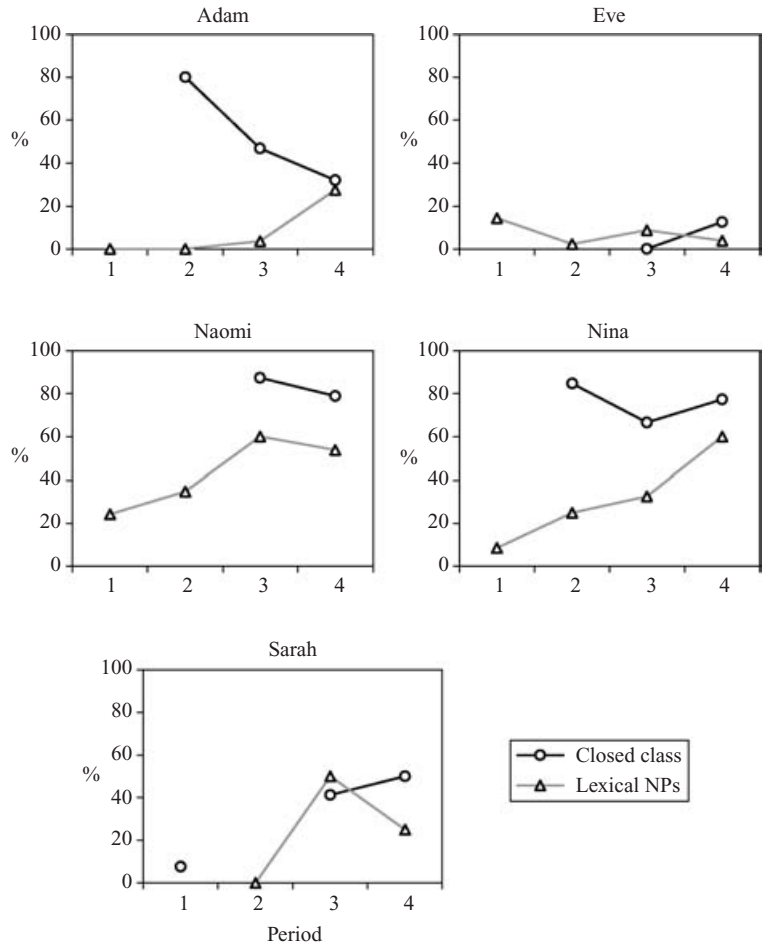

Fig. 3. Present progressive auxiliary marking with third person closed-class and lexical NP subjects by time.

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

general, by chunking it together with the subject. One child, Eve, clearly does not display this pattern. Eve produced both copulas and auxiliaries quite infrequently (in 18% and 6% of obligatory contexts respectively). It is plausible that she had not (at least in the time period covered by the transcripts) followed the strategy that it is argued the other children have followed, of internalizing subject + *be* chunks. This could lead both to her low rate of *be* in general, as well as to the lack of any difference between open- and closed-class subjects.

The almost complete lack of any copulas or auxiliaries with open-class subjects in the first half of the data from three children poses a challenge to full competence theories, since it suggests that these children did not in fact have any genuinely productive means of producing the copula or the auxiliary. The patterns observed here are also problematic for structure building models, since the children do produce a substantial amount of inflectional material, so they must be said to have acquired IP, making their limitations hard to account for.

Variation among individual closed-class subjects

Constructions, under the present view, range from the very general to the very specific. It was argued earlier that copular constructions, present progressives, and clauses involving 3sg agreement are separate constructions, relying on pieces of knowledge which are initially distinct from one other, accounting for the fact that they are learned at different rates by different children. These are quite general construction types. However, the hypothesis is that general constructions are built up out of more specific ones: a schema such as $NP_{subj\text{-}3sg}$'s V-*ing* depends on the prior existence of more lexically specific schemas such as *he's* V-*ing* and *it's* V-*ing*. In this section, evidence is presented that early production of copula and auxiliary *be* is in fact dependent upon highly lexically specific constructions of this sort.

To test whether such schemas are involved in the production of *be*, usage of *be* was tabulated according to the particular lexical item occurring as subject. Thirteen closed-class subjects were considered: *I*, *you*, *he*, *she*, *it*, *we*, *they*, *this*, *that*, *these*, *those*, *here* and *there*. The percent correct in obligatory contexts for each of the subjects was compared to a baseline, which was the percent correct for the remaining 12 combined (Fisher's Exact Test, two-tailed). Lexical subjects were not used in these comparisons, so these results are distinct from those presented in the previous section. The data for copular constructions are shown in Table 5 and Figure 4. The variation from subject to subject is striking. Adam has 9 closed-class subjects which differ significantly from the baseline, Sarah has 6, Nina has 5, Eve has 3 and Naomi has 1.[3] No correction

[3] The phenomenon of non-nominative subjects (e.g. *me go*) could potentially be relevant in this kind of analysis, since some accounts (e.g. Schütze, 1997) predict an interaction

TABLE 5. *Variation among individual closed-class subjects in copular constructions*

	Adam		Eve		Naomi		Nina		Sarah	
Pronoun	Abs Corr	Pres Sig	Abs Corr	Pres Sig	Abs Corr	Pres Sig	Abs Corr	Pres Sig	Abs Corr	Pres Sig
I	99	20	16	5	2	17	11	3	18	22
	17%	<***	24%	>	89%	>	21%	<***	55%	<
you	21	5	11	1	2	4	4	3	7	1
	19%	<**	8%	<	67%	<	43%	<**	13%	<*
he	12	2	16	3	1	2	10	84	32	11
	14%	<*	16%	<	67%	<	89%	>	26%	<***
she	3	0	1	0	0	1	0	1	5	11
	0%	<	0%	<	100%	>	100%	>	69%	>
it	10	113	58	21	3	37	2	91	8	40
	92%	>***	27%	>	93%	>	98%	>***	83%	>***
we	5	1	0	0	0	0	0	0	1	1
	17%	<	–	=	–	=	–	=	50%	<
they	18	11	8	5	0	0	1	25	2	2
	38%	<	38%	>	–	=	96%	>	50%	<
this	37	100	8	27	9	36	27	25	35	5
	73%	>***	77%	>***	80%	<	48%	<***	13%	<***
that	298	161	298	28	6	98	53	523	47	86
	35%	<***	9%	<***	94%	>*	91%	>***	65%	>*
these	1	23	2	0	0	0	1	3	0	3
	96%	>***	0%	<	–	=	75%	<	100%	>
those	1	35	5	0	0	7	9	32	5	4
	97%	>***	0%	<	100%	>	78%	<	44%	<
here	31	11	4	6	5	16	16	139	21	29
	26%	<**	60%	>**	76%	<	90%	>	58%	>
there	17	13	28	12	5	41	7	44	9	34
	43%	<	30%	>	89%	>	86%	<	79%	>**
All	553	495	455	108	33	259	141	973	190	249
	47%		19%		89%		87%		57%	

For each closed-class subject, there is for each child a 2-by-2 set of values showing the number of times the copula was absent with that subject, the number of times it was present, the percentage of the time it was present, and whether that percentage differed significantly from the average percentage of all other closed-class subjects for that child.

has been made for multiple comparisons, so it is likely that one or two of these are false positives, but there is no doubt that there is a great deal of variation in the rates of copula marking from subject to subject.

This suggests that whether the copula is produced often depends to a considerable extent on whether the child has a construction which includes the necessary closed-class subject along with its copula form. Consider, for

between subject case and agreement. However, Nina was the only child to produce more than a few non-nominative subjects, and only two were at all common: *my* and *her*, so this does not appear to be a factor in the results presented here.

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

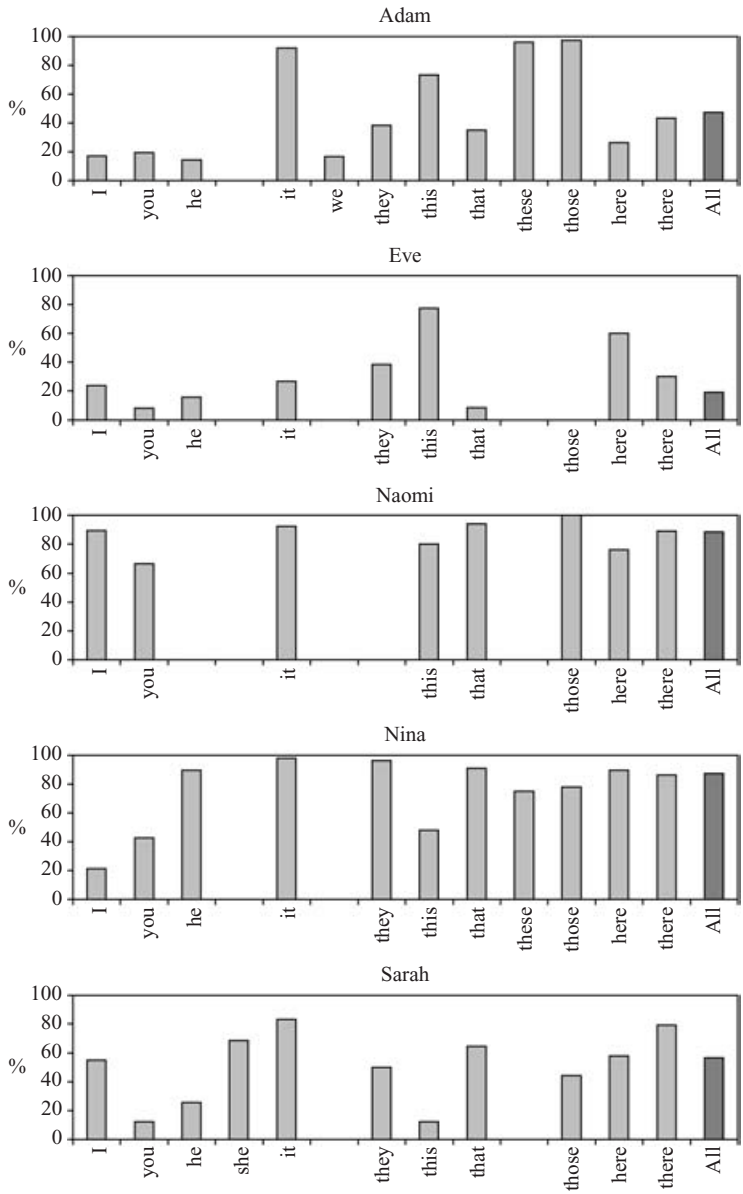

Fig. 4. Variation among individual closed-class subjects in copular constructions. Subjects are only shown on the *x*-axis when they occurred in at least four obligatory contexts.

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

a particularly clear example, Adam's production of the copula with the subjects *those* and *you*. He produced the copula correctly with *those* 35 out of 36 times (see examples in (8)), whereas he produced the copula only 5 out of 26 times with *you* as subject (examples in (9)).

(8) (*a*) Dose are circus men. (Adam, 3;0) (35 examples like this)
(*b*) Dose mixer cakes. (Adam, 3;4) (the sole example like this)

(9) (*a*) You are strawberry. (Adam, 2;10) (5 examples like this)
(*b*) You very sticky. (Adam, 3;0) (21 examples like this)

This can be explained by assuming that Adam had extracted a construction of the form *those are* XP_{pred}, but a similar *you are* XP_{pred} construction was slower to develop. Because Adam's *you are* XP_{pred} construction was relatively slow to develop, most second person copular sentences had to be produced with a more general but immature and 'pared down' NP_{subj} XP_{pred} construction, which lacks an overt copula.

These results would not easily be accommodated within a full competence or structure building model in which production of correctly agreeing copulas is taken as evidence for the child having knowledge of the relevant principles of inflection. Such theories would have difficulty explaining, for example, why Adam, who produced the copula very reliably with *it*, *this*, *these* and *those*, produced it so much less frequently with *I*, *you*, *he*, *that* and *here*. The fact that Adam produced the copula so reliably with *those* suggests that whatever putative grammatical or pragmatic factors (e.g. underspecification of functional heads) cause or permit the omission of inflection were not operative at the point he was producing these utterances. It appears that the only way then to explain the missing copulas with *you* is to appeal to lack of knowledge of subject + copula constructions like *those are*.

We can rule out that it is the allomorphs of *be*, rather than subject + copula units, that are learned one by one. It can be seen in Table 5 and Figure 4 that subjects which require the same allomorph of *be* can nonetheless vary greatly in the proportion of the time that *be* is overt. Returning to the examples in (8) and (9), a full competence account has to attribute to Adam knowledge of the allomorph *are*, based on his correct productions of *those are*. He also clearly knows the pronoun *you*, and so the relative rareness of *you're* or *you are* cannot be explained as resulting from incomplete lexical learning, at least not from incomplete learning of what are normally taken to be lexical items. It may be possible to claim that although Adam knows the allomorph *are*, he does not know that it is the allomorph required with *you*. However, four of the children (the exception being Naomi) show substantial variation even among the various 3sg subjects (e.g. *he*, *it*, *this*, *that*, etc.). For instance, it can be seen in Table 5 that Adam and Eve both produced the copula much more reliably with *this* than with *that*, whereas Sarah and Nina show exactly the opposite pattern.

To capture this variation, it would be necessary to claim that the children have not learnt that *-'s* is the 3sg allomorph, implying in effect that they have no knowledge of the formal property of agreement, which would seem to be at odds with any full competence hypothesis.

The variation which can be observed from child to child suggests also that an explanation of these asymmetries in grammatical terms would be unlikely to succeed. Rather, the randomness observed suggests that what we are looking at is in fact a form of lexical learning, but that it is constructions with open slots which are being learned, not just lexical items to be inserted into the terminal nodes of phrase markers.

One further observation to make about these results is that even Eve, for whom earlier no significant differences in the overtness of *be* with closed- vs. open-class subjects were found, shows at least two closed-class subjects which seem to have been learned along with copula *be*: *this is* and *here* (*i*)*s*. This suggests that it is not that Eve was not using the construction-based strategy at all for learning *be*, but rather that she was doing so to a lesser extent than the other children, at least during the time period for which data are available.

A possible confound worth ruling out arises from the fact that the figures presented above are summed over time. If a child only started using a particular closed-class subject relatively late, after inflection has begun to be used more consistently across the board, then that could explain why *be* is found to occur more frequently with that closed-class subject. However, this appears not to be the case. Considerations of space preclude a full presentation of the data, but Table 6 provides a brief summary. It shows, for each child, how many closed-class subjects differed significantly from the expected rate of *be* occurrence for at least one of the four arbitrary time periods. This is compared to the number of closed-class subjects for which this was true on the basis of all the data. As can be seen, the figures are very similar, suggesting that the observed asymmetries hold at specific points in time as well as over the data as a whole.

Turning now to auxiliary *be*, similar patterns were found, though they are less striking. These data are shown in Table 7 and Figure 5. Adam and Nina each have 3 subjects which differ significantly from the baseline for auxiliary *be* (defined again as the average over all other closed-class subjects), Eve and Naomi have 1 each, and Sarah has none. There is clearly a considerable amount of variation, though less than was observed with copula *be*. This probably reflects the fact that the children tended to use a smaller range of closed-class subjects with auxiliary *be*, and also the fact that present progressives are less common than copular constructions in the corpora, so there is less data.

It must be noted that Adam's abnormally frequent use of both copulas and auxiliaries with *it* subjects is in fact spurious. As Brown (1973) observed, Adam clearly misanalysed *it's*: he used it consistently both in environments requiring *it's* and those requiring *it* (e.g. *it's hurts*, 2;11). This is a genuine segmentation error, and seems to be relatively rare.

TABLE 6. *Variation among closed-class subjects in copular constructions, controlled for time*

	Adam		Eve		Naomi		Nina		Sarah	
Condition	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$
By time	11	12	2	9	1	7	6	10	6	9
All	9	11	3	9	1	7	5	10	6	10

The 'Sig' column shows, for each child, how many of the 13 closed-class subject types showed copula frequencies significantly different from all other closed-class subjects at $p \leqslant 0{\cdot}05$. The '$N \geqslant 7$' column shows how many of the 13 subjects occurred in obligatory contexts for the copula 7 or more times, or showed copula frequencies significantly different from the baseline, giving a very rough indication of how many COULD have differed significantly from the mean of all closed-class subjects. The 'By time' row reports whether the conditions just discussed were met for any of the four quarters of the data. The 'All' row summarizes the information in Table 5.

TABLE 7. *Variation among closed-class subjects in present progressive constructions*

	Adam		Eve		Naomi		Nina		Sarah	
Pronoun	Abs Corr	Pres Sig	Abs Corr	Pres Sig	Abs Corr	Pres Sig	Abs Corr	Pres Sig	Abs Corr	Pres Sig
I	525	26	157	8	50	89	58	6	38	14
	5%	<***	5%	<	64%	<*	9%	<***	27%	<
you	20	4	17	0	1	1	7	3	5	0
	17%	>	0%	<	50%	<	30%	<	0%	<
he	49	7	30	0	1	5	23	102	19	5
	13%	<	0%	<	83%	>	82%	>***	21%	<
she	10	0	5	0	1	1	0	2	6	4
	0%	<	0%	<	50%	<	100%	>	40%	>
it	2	72	7	1	0	8	2	10	3	3
	97%	>***	13%	>	100%	>	83%	>	50%	>
we	82	5	9	2	0	3	7	1	2	2
	6%	<*	18%	>	100%	>	13%	<*	50%	>
they	48	14	3	4	3	10	12	14	0	0
	23%	>	57%	>***	77%	>	54%	<	—	=
this	12	2	2	0	0	0	0	0	2	0
	14%	<	0%	<	—	=	—	=	0%	<
that	18	1	1	0	0	0	1	6	0	1
	5%	<	0%	<	—	=	86%	>	100%	>
these	0	0	0	0	0	0	0	0	0	0
	—	=	—	=	—	=	—	=	—	=
those	0	1	0	0	0	0	1	0	0	0
	100%	>	—	=	—	=	0%	<	—	=
All	766	132	231	15	56	117	111	144	75	29
	15%		6%		68%		56%		28%	

See Table 5 for explanation, though this table reports on auxiliary *be* rather than copula *be*.

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

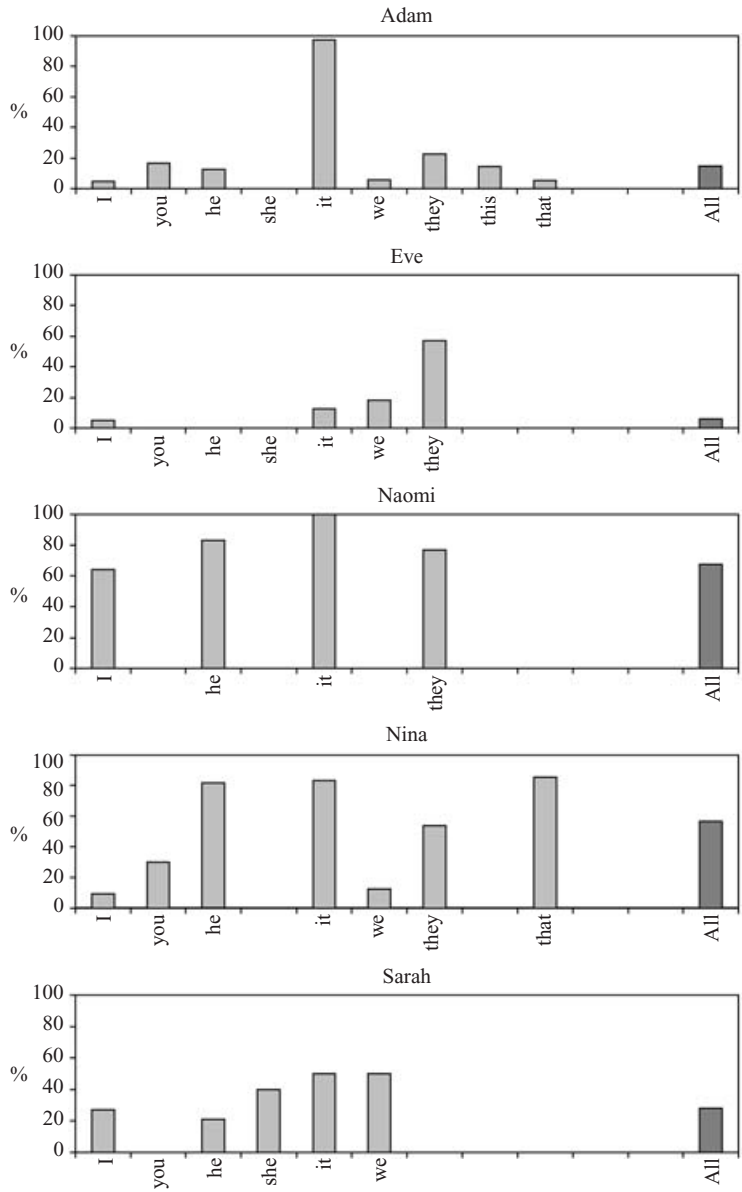

Fig. 5. Variation among individual closed-class subjects in present progressive constructions.

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

TABLE 8. *Variation among closed-class subjects in progressive constructions, controlled for time*

	Adam		Eve		Naomi		Nina		Sarah	
Condition	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$
By time	4	9	1	5	0	2	2	4	0	3
All	3	9	1	6	1	3	3	7	0	3

See Table 6 for explanation, though this table reports on auxiliary *be* rather than copula *be*.

The asymmetries observed for auxiliary *be* cannot be explained as reflecting the summation of data over time, as Table 8 shows. Most of the subjects which differ significantly from the norm in terms of how often they occurred with auxiliary *be* also do so for at least one of the four more limited time periods.

Brown (1973) discusses results related to those presented in this section, though the tables he presents (pp. 267–8) document only the differences between allomorphs, not the differences between particular closed-class subjects. Kuczaj (1985) observes that '[i]n the acquisition of a *be* allomorph, the predominant pattern was one in which children learn to use the allomorph in quite specific contexts' (Kuczaj, 1985: 113). For instance, his son Abe 'sometimes used copula *are* in declarative sentences beginning with *those* or *these* … but consistently omitted [it] in all other obligatory contexts' (p. 113). This is very similar to the patterns observed here, and suggests that these results are robust. Interestingly, Santelmann & Jusczyk (1998) and Tincoff, Santelmann & Jusczyk (2000) report similar results regarding the comprehension abilities of children at 1;6. Their experiments, using the headturn preference procedure, showed that infants recognize the co-occurrence relationships between *is* and *-ing* and *was* and *-ing*, but not between *are* and *-ing* or *were* and *-ing*.

Controlling for predicate type

A very different theory attempting to explain variation in the frequency of copula marking as a function of predicate type has been proposed by Becker (2000). It needs to be considered whether her proposal could provide an alternative account for some of the results reported for copula *be* in the previous sections. As discussed earlier, the empirical predictions of Becker's theory are that *be* should be overt with nominal predicates (10*a*), sometimes overt and sometimes not with adjectival predicates (10*b*), and should be absent with locative predicates (10*c*).

(10) (*a*) He is/*Ø a dog.
(*b*) She is/Ø happy.
(*c*) It *is/Ø in the kitchen.

TABLE 9. *Copula marking with nominal, adjectival and locative predicates*

	Nominal			Adjectival			Locative			Significance		
Child	Abs	Pres	%	Abs	Pres	%	Abs	Pres	%	Nm/Ad	Nm/Lc	Ad/Lc
Adam												
CC	292	278	49	131	107	45	8	12	60	>	<	<
Lex	6	9	60	30	9	23	46	5	10	>*	>***	>
Total	298	287	49	161	116	42	54	17	24	>*	>***	>**
Eve												
CC	287	59	17	49	9	16	10	8	44	>	<**	<*
Lex	16	7	30	22	3	12	18	2	10	>	>	>
Total	303	66	18	72	12	14	28	10	26	>	<	<
Naomi												
CC	8	119	94	9	51	85	0	4	100	>	<	<
Lex	1	0	0	18	12	40	11	9	45	<	<	<
Total	9	119	93	31	63	67	11	13	54	>***	>***	>
Nina												
CC	69	565	89	24	96	80	17	59	78	>**	>**	>
Lex	4	12	75	19	22	54	73	36	33	>	>**	>*
Total	76	577	88	61	122	67	99	96	49	>***	>***	>***
Sarah												
CC	75	102	58	63	41	39	3	11	79	>**	<	<**
Lex	6	3	33	8	2	20	5	0	0	>	>	>
Total	81	105	56	71	43	38	8	11	58	>**	<	<

CC, closed-class; Lex, lexical NP; Nm, nominal; Ad, adjectival; Lc, locative.

If the distribution of closed-class and lexical NP subjects is skewed with respect to these classes, then this could potentially provide an alternative explanation for the asymmetries observed so far. In fact, there is such a skewed distribution, as Becker shows. Most notably, nominal predicates tend to have closed-class subjects. This appears to be simply a fact about discourse (see Clancy 2000 for some relevant discussion): sentences such as (11*a*), with a closed-class subject, are much more frequent than those such as (11*b*), with an open-class subject.

(11) (*a*) That's Rudolph. (Sarah, 3;1)
(*b*) Colleen's mommy's a big girl. (Nina, 2;3)

First some counts are reported related to those carried out by Becker (2000). The relevant figures are reported in Table 9 and displayed in Figure 6. In these figures, all closed-class subjects are included, not just third person ones, in order to include as much data as possible. Significance was calculated using Fisher's Exact Test, one-tailed to test Becker's predictions, but two-tailed for those results which went against her predictions. The files used in this study overlap only partly with those used by Becker (2000) (*cf.* Table 1

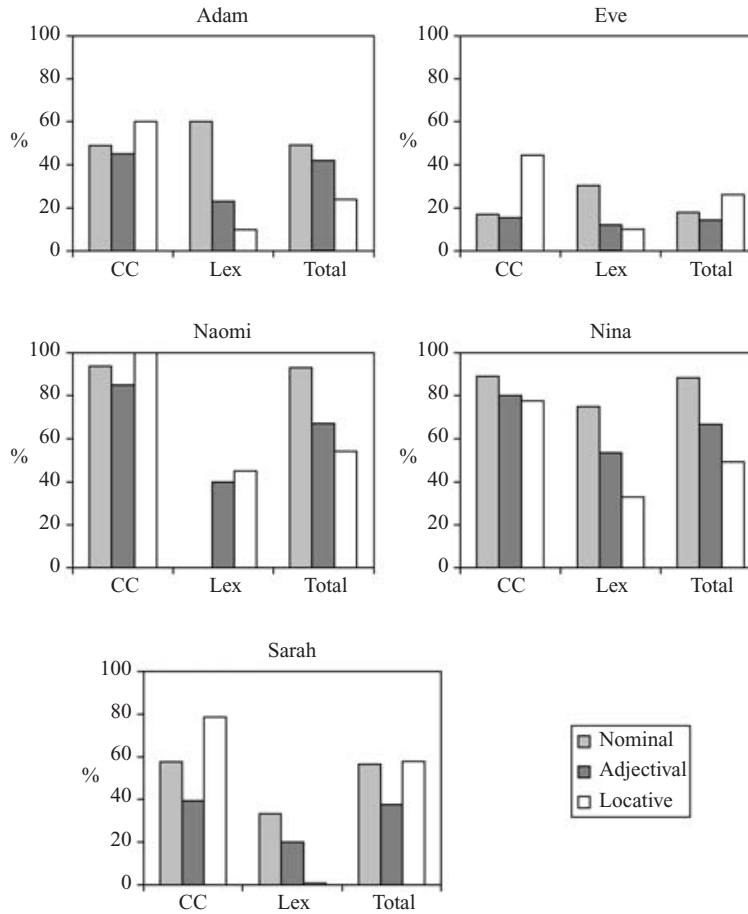

Fig. 6. Copula marking with nominal, adjectival and locative predicates. CC, closed-class; Lex, lexical NP subjects.

with her Table 3.1 on p. 85), so this is not intended as a replication of her counts.

Considering only the totals, three of the five children – Adam, Naomi and Nina – appear to show support for the asymmetries Becker claims. The comparisons among Eve's construction types are nonsignificant despite a reasonable amount of data, a fact also noted by Becker. Sarah shows the expected greater proportion of overt copulas with nominals compared to adjectives, but she also shows more overt copulas with locatives compared to adjectives and nominals, which is the opposite of what Becker predicts, though

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

these comparisons are not significant. Three of the five children, then, seem to offer support for Becker's empirical claims, while two pose problems.

However, in Table 9 and Figure 6 the data have also been divided according to open- vs. closed-class subjects. Since Becker claims this is not an important factor (p. 150), her results should continue to hold in both cases. They do continue to hold for Nina. But for Adam, although the predicted asymmetries can be observed with open-class subjects, the figures for closed-class subjects are virtually the same across the three predicate types. And for Naomi, once the data are divided up like this, there are no longer any significant differences. It appears that the observed high rate of copulas with nominal predicates that we see in the total figures for Naomi probably reflects the fact that all but one of her nominal predicates had closed-class subjects. So, while some children show some of the effects of predicate type that Becker predicts, not all children do, and for some, like Naomi, the real effect might actually be due to open- vs. closed-class subjects. Furthermore, once the data are divided, some of the comparisons for Eve and Sarah are actually significant in the opposite direction to that predicted by Becker.

More importantly for the purposes of this study, it is necessary to examine whether the asymmetries observed earlier continue to hold when predicate type is held constant. This inevitably entails a loss of statistical power, since there are obviously far fewer sentences involving any one type of predicate than there are copular constructions in total. Table 10 and Figure 7 present comparisons of copula marking with open- and closed-class subjects, broken down by predicate type. These figures, unlike those in Table 9 and Figure 6, include only third person closed-class subjects, for reasons discussed earlier. Of the 15 comparisons, 7 are significant, and they all go in the direction predicted by the constructivist account: copulas are produced more reliably with closed-class subjects. Except for Eve, who as already noted does not appear to use subject+*be* units to a great extent, the failure to obtain significance in the other comparisons can plausibly be explained in many cases as being due to there being relatively little data.

Becker (2000) actually argues against the importance of the open/closed-class asymmetry, claiming that 'Adam's and Naomi's data illustrate conclusively that the overtness of the copula is not dependent on the presence of a pronominal subject' (p. 150). However, she counted only 7 of Adam's files, whereas this study is based on all of Adam's first 30 files, and it turns out that for both adjectival and locative predicates, Adam did use significantly more copulas with closed-class subjects ($p \leqslant 0{\cdot}001$ in both cases). It is also worth pointing out that for nominal predicates, where no significant difference was found, all of Adam's 9 overt copulas with lexical NPs occured in the last quarter of the data. As for Naomi, the only predicate type for which both closed- and open-class subjects are represented is adjectival predicates, for which closed-class subjects do show overt copulas more frequently

TABLE 10. *Copula marking with closed-class and lexical NP subjects, by predicate type*

	Closed-class (3sg/3pl)			Lexical NP subjects			
Child	Absent	Present	%	Absent	Present	%	Significance
Adam							
Nom	236	268	53	6	9	60	<
Adj	69	94	58	30	9	23	>***
Loc	4	11	73	46	5	10	>***
Eve							
Nom	271	56	17	16	7	30	<
Adj	46	7	13	22	3	12	>
Loc	8	7	47	18	2	10	>*
Naomi							
Nom	8	119	94	1	0	0	>
Adj	6	37	86	18	12	40	>***
Loc	0	1	100	11	9	45	>
Nina							
Nom	68	561	89	4	12	75	>
Adj	13	94	88	19	22	54	>***
Loc	14	59	81	73	36	33	>***
Sarah							
Nom	66	87	57	6	3	33	>
Adj	47	35	43	8	2	20	>
Loc	3	8	73	5	0	0	>*

($p \leqslant 0{\cdot}001$). So this claim must be rejected. Furthermore, even if there are some children for whom the asymmetry does not hold, this would not constitute an argument that the asymmetry is not important.

It can also be noted that the asymmetry between open- and closed-class subjects found with auxiliary *be* in the present progressive, shown above in Table 4 and Figure 3, provides further evidence that the open/closed-class subject distinction is a key factor in the overtness of *be*. Predicate type cannot be a confound here because only a single predicate type – the present progressive – is involved.

Finally, it needs to be checked that the asymmetries observed among individual closed-class subjects continue to hold when predicate type is controlled. This involves a very large number of comparisons, so only a summary is provided in Table 11, showing for each predicate type how many closed-class subjects differed significantly from the norm for that predicate type. Considering that we are now dealing with even smaller numbers (i.e. copular constructions with a particular subject and a particular predicate type) the results are surprisingly robust, especially for nominal predicates, which often have closed-class subjects. It appears that the varying rates of *be* observed with

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

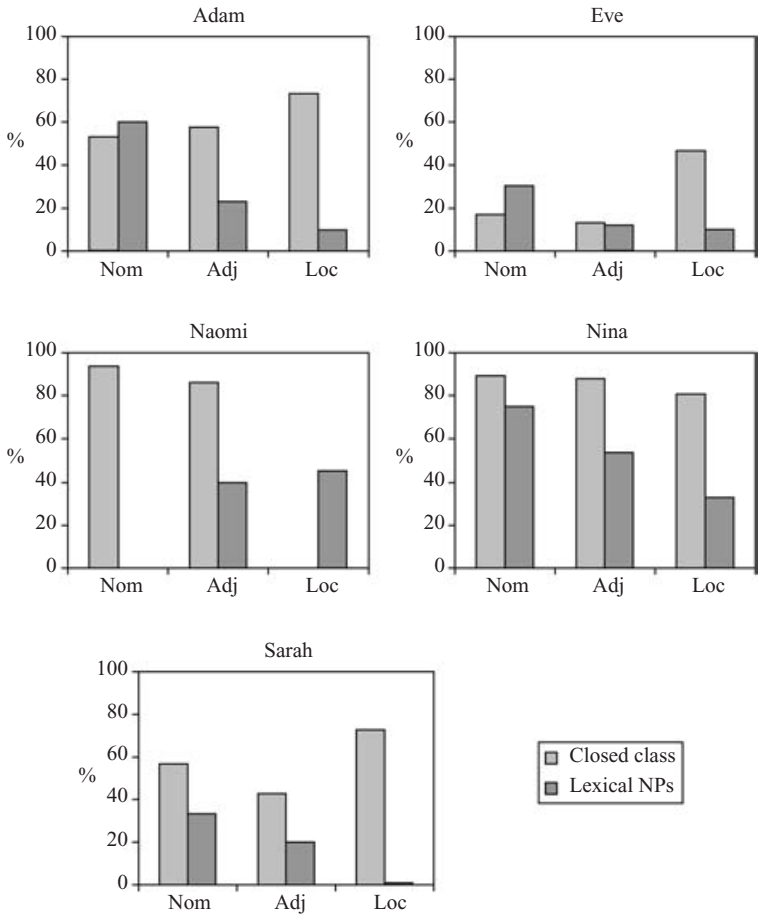

Fig. 7. Copula marking with third person closed-class and lexical NP subjects, by predicate type. Nom: nominal; Adj, adjectival; Loc, locative.

different subjects do not follow from those different subjects occurring with different types of predicate.

In sum, the results obtained in previous sections are not greatly compromised when predicate type is controlled for, although there is some loss of statistical power. Furthermore, the results call into question the empirical basis for Becker's (2000) specific proposals about predicate type as a major determinant of copula overtness. While predicate type probably does play a role, more research is needed to determine which patterns, if any, are consistent across children.

https://doi.org/10.1017/S0305000902005512 Published online by Cambridge University Press

TABLE 11. *Variation among closed-class subjects in different types of copular constructions*

	Adam		Eve		Naomi		Nina		Sarah	
Condition	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$	Sig	$N \geqslant 7$
Nom only	8	9	3	6	1	3	4	6	2	6
Adj only	3	6	0	3	0	4	3	5	3	5
Loc only	1	2	0	1	0	0	4	4	1	2
All	9	11	3	9	1	7	5	10	6	10

See Table 6 for explanation of the 'Sig' and '$N \geqslant 7$' columns. This table shows how many closed-class subjects differed significantly from the others when copular constructions were divided into nominal, adjectival and locative predicates.

GENERAL DISCUSSION

At the outset of this study, two empirical questions were posed, repeated below, which were designed to investigate whether inflection emerges as a unitary category, consistent with recent generative accounts, or whether its development is more piecemeal and initially lexically restricted, as would be expected under a constructivist account.

(I) What are the relative rates of acquisition of copula *be*, auxiliary *be* and 3sg present agreement *-s*?

(II) What, if any, is the effect of the subject (whether it is open- or closed-class, and which closed-class pronoun or demonstrative it is) on the realization of copula and auxiliary *be*?

With regard to the first question, it was found that the relative rates of acquisition of copula *be*, auxiliary *be* and 3sg present agreement *-s* vary significantly from child to child. Some children show large differences between the rates at which they supplied these three morphemes, the most striking example being Nina, who used copula *be* correctly 79% of the time, but 3sg agreement only 13% of the time. This is not predicted by full competence theories such as Schütze (1997) or Rice *et al*. (1998), where the presence or absence of all three of these morphemes is held to depend on the same underlying feature(s). The data are also inconsistent with structure building models such as Radford (1990), where it is claimed that IP develops at some particular point in time, whereafter all inflections are expected to be present, and problematic for the version in Radford (1996) which relies only upon an unspecified truncation mechanism to account for missing inflections. Furthermore, although all five children correctly supplied copula *be* more frequently than auxiliary *be*, there were no consistent patterns with the relative rates of copula *be* compared to 3sg agreement, or auxiliary *be* compared to 3sg agreement. This creates difficulties for theories such as Hyams (1999) and Becker (2000), where factors such

as modality or aspect could play a role in determining omissions, since presumably the constructions have the same aspectual and modal properties for each of the children, yet there are still substantial individual differences in the relative rates at which inflections are supplied in each construction. The patterns observed support a constructivist account in which the correct production of these three morphemes depends upon distinct schemas for each construction, so that the constructions are independent and may develop at different rates.

It would also be possible to propose three (or more) independent factors within a full competence or structure building account. Recent work in theoretical syntax suggests that there may be many more functional heads than has previously been assumed (Cinque, 1999). If these could be independently underspecified in the same way that Schütze (1997) assumes Tns and Agr can be, then the observed asymmetries could be accommodated. Such a theory would be difficult to distinguish empirically from the constructivist account with respect to this particular question.

The second question which was asked was whether the subject has any effect on the realization of copula and auxiliary *be*. The results indicated that third person closed-class subjects were significantly more likely than lexical NP subjects to show overt copula and auxiliary *be*. This is consistent with the constructivist account, because forms such as *he's* and *that's* can be learned as chunks, but it is less likely that open-class subjects, which are less frequent, would be learned as chunks along with copulas or auxiliaries which may sometimes follow them. This result appears to hold up among individual semantic classes of copular constructions, so predicate type, while it may be a factor, is not sufficient in itself to explain the findings (*cf.* Becker, 2000).

This result may be able to be predicted via an entirely distinct line of argument within a theory such as Schütze's (1997) Agr/Tns Omission Model.[4] It is well known that children learning English produce substantial numbers of null subjects. Schütze (1997) claims that the omission of subjects is licensed when Tns is underspecified. It seems reasonable to assume that subject omission also depends upon pragmatic factors, such that pronouns, which are more recoverable, are omitted more often than open-class NP subjects, which are less recoverable. The underspecification of Tns also results in the omission of *be*. So it could be that *be* is omitted equally often in sentences with open- and closed-class subjects, but that in the latter, the omission of Tns, along with the recoverability of the subject, licenses subject drop as well as *be* drop, at which point an uncountable utterance like *green* would result. Thus, according to this line of reasoning, many cases where *be* has been omitted with closed-class

[4] I thank Ken Wexler for a suggestion similar to this, which I have modified somewhat here.

subjects are simply not being counted. A potential problem such an account would face is that it is not clear whether Schütze's (1997) claim that null subjects are licenced by underspecified Tns is correct, because children also omit subjects of clauses showing 3sg agreement, which must be specified for Tns (Sano & Hyams, 1994; Schütze & Wexler, 2000). Naturally this does not rule out underspecified Tns as a factor in subject omission, but more evidence is needed to establish a correlation.

However, one aspect of the data on open- vs. closed-class subjects strongly supports the idea that early copula production depends on lexically specific frames. This is the fact that three of the five children produced between them only 1 copula and 2 auxiliaries with lexical NP subjects in the first half of their data. This suggests that these children had no means of producing copula and auxiliary *be* that was truly productive in the sense that it could generalize across subjects.

Question (II) also asked whether there are differences in the rate of *be* among particular closed-class subjects. The results show that this is indeed the case, especially for copular constructions. This is expected under the constructivist account, in which early constructions revolve around particular lexical items, forming the raw material out of which more general constructions can be abstracted. However, the asymmetries pose a challenge to theories which assume that the child has a category such as IP, which would be expected to result in inflection being productive. The factors which have been proposed to be responsible for the omission of inflectional material, such as the underspecification of functional heads, do not make reference to which particular closed-class subject a clause contains.

Significant variation was found across children in terms of which particular closed-class subjects frequently occurred with overt *be* and which did not, suggesting that grammatical factors are unlikely to systematically account for the asymmetries observed. Rather, it appears that to a large extent what determines whether a child uses *be* is whether the child has a subject + *be* construction for the subject they wish to use. For instance, the fact that Adam used the copula *are* correctly with *those* 35 out of 36 times suggests that there is no factor, grammatical, pragmatic, or otherwise, which licenses non-finite clauses in the speech of this child at this stage. Rather, it is proposed that the only reason Adam often omits *be* with most other subject types at this stage is that he lacks the requisite constructions which would allow him to produce it. Similar results can be observed for some of the other children.

It may be possible to import this claim into a full competence theory. This would require the syntactic component to 'look ahead' to the lexicon to see whether the necessary unit(s) were known, in order to determine whether features leading to overt inflection could be specified. An interesting approach along these lines, though not addressing the issue of chunked lexical items, is pursued by Phillips (1995). However, if a full competence theory relying on

incomplete lexical and/or morphological knowledge to account for missing inflections is to capture the range of asymmetries documented in this paper, it seems probable that the kind of lexical and/or morphological knowledge that would need to be postulated would likely be very similar to the lexically specific constructions which are assumed here to underlie early productions. Although such a theory could accommodate the empirical facts, the problem is simply that it is hard to see what role the assumption of the child having the full clausal architecture would play in such an account, and how it could be justified. See Pizzuto & Caselli (1992, 1993) for related arguments, and Hyams (1992*b*) for a defence of the nativist position.

In sum, the data presented in this study argue against full competence hypotheses in two ways. Firstly, the evidence that many if not all early uses of inflection are tied to lexically specific frames shows that the production of correctly agreeing forms cannot be taken as evidence for the child knowing the relevant morphemes and principles of agreement. This has been the strongest argument for attributing full competence to children, but the data show that their abilities to control the syntactic process of agreement are much more limited than they appear at first sight. Secondly, the mechanisms which have been proposed within full competence theories to account for missing inflections are too monolithic to capture the actual developmental patterns which can be observed.

With respect to structure building models, the data offer strong support for the idea that children do indeed build up functional structure in the course of development. However, it is clear that there is no particular point at which a category such as IP is triggered, and such models will have to propose more detailed accounts of exactly how functional categories are acquired if they are to capture the detail of the developmental patterns.

The present study has attempted to make some inroads into this kind of investigation by proposing explicit constructions to underlie children's production of particular inflectional morphemes, and by discussing the kinds of processes which could be involved in the abstraction of these constructions. Two features of this model appear to be crucial in accounting for the data: children's knowledge of grammar is best expressed in terms of relatively limited positive constructs which license certain classes of utterances, rather than in terms of a more general grammar subject to negative constraints (e.g. underspecification of heads, or a truncation mechanism); and secondly, constructions are like lexical items in that there is a certain amount of randomness in terms of when they are acquired by particular children.

If forms such as *he's* and *that's* are units, this may have implications for some other topics in the acquisition of English which have been discussed in the recent literature. Schütze & Wexler (1996) and Schütze (1997) observe a correlation between agreement and correct use of nominative case pronouns, which they argue supports the idea that children have innate knowledge of

principles of case assignment. However, this correlation could alternatively be due, at least in part, to children using large numbers of chunked forms such as *he's* and *I'm* which instantiate both agreement and nominative case. This suggestion was made by Rispoli (1999), and this study has provided evidence that forms such as *he's* and *I'm* do indeed function as units for children at the relevant stage of development.

Another fact which may be explicable under the present approach is that null subjects are found with main verbs marked for 3sg present agreement much more frequently than they are found with overt copula/auxiliary *be* (Sano & Hyams, 1994; Schütze & Wexler, 2000). If *be* is produced primarily in constructions in which it is chunked together with a subject, then it follows that null subjects could not be found with *be*. But null subjects would still be expected to occur with agreeing main verbs, since there the subject presumably does not form a unit with the agreement morpheme.

However, while these are interesting possibilities for future research, the development of accounts such as these clearly depends upon having a well-articulated construction-based theory of learning. Further empirical studies will be necessary to determine exactly what range of constructions is necessary and sufficient to account for what children can produce at various stages of development, in various languages. It will then become possible to be more explicit about what kind of heuristics children use in extracting units from the input, in breaking them down, in generalizing them, and so on. Promising initial steps towards addressing these kinds of issues have been made by researchers such as Peters (1983), Tomasello & Brooks (1998) and Tomasello (2000), but there is still a clear need for further theoretical development.

Another essential claim of the present account which must be worked out in more detail is that children derive 'pared down' versions of the constructions to which they are exposed. It is necessary to determine what factors influence which items are left out. As mentioned earlier, relative uninformativeness, semantic complexity and lack of acoustic salience could be three possible candidates. There has been a substantial amount of research on this topic, but the issues are still far from settled. This is a crucial question to address if a constructivist theory is to provide a good account for cross-linguistic similarities and differences in developing grammars, since a central fact about children's speech is that they omit various types of words and morphemes, but the challenge is to explain why certain items are omitted in some languages but not in others, and so on. The results presented in this study suggest that frequencies of construction types in the input interact in complex ways with the factors which make certain morphemes liable to be omitted. Specifically, although copulas and auxiliaries are in general slow to be mastered productively, they are used far more reliably when they occur as components of highly frequent subject + *be* constructions. Finally, it is also worth noting that the notion of construction assumed in this study relies quite heavily

upon word order, which may not be appropriate in dealing with languages with freer word order. Much work remains to be done in this regard. See Rowland & Pine (2000) for more discussion of the need for a more explicit construction-based theory of learning. Tomasello (2000) argues that an equally large problem looms for generative theorists: explaining how the input is linked up to the putative innate structures.

CONCLUSION

This study has documented several kinds of asymmetries in the acquisition of inflection in English, arguing that they are problematic for recent generative accounts where inflection is held to emerge as a unitary category. It has been argued that the data are more compatible with the view that initial knowledge of inflection takes the form of constructions of varying levels of specificity, which are only gradually integrated into more abstract constructions. The data reported in this paper hopefully demonstrate that the development of more explicit construction-based theories of learning is a goal well worth pursuing.

REFERENCES

Aldridge, M. (1989). *The acquisition of INFL*. Bloomington, IN: Indiana University Linguistics Club Publications.

Bates, E. & Goodman, J. C. (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (ed.), *The emergence of language*. Mahwah, NJ: Erlbaum.

Becker, M. (2000). *The development of the copula in child English: the lightness of* be. Unpublished doctoral dissertation, University of California, Los Angeles.

Braine, M. D. S. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development* **41**, serial no. 164.

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

Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child Development* **39**, 433–48.

Cinque, G. (1999). *Adverbs and functional heads: a cross-linguistic perspective*. New York: Oxford University Press.

Clahsen, H., Eisenbeiss, S. & Penke, M. (1996). Lexical learning in early syntactic development. In H. Clahsen (ed.), *Generative perspectives on language acquisition*. Amsterdam: John Benjamins.

Clancy, P. (2000). Transitivity in Korean acquisition: discourse-structural foundations. In E. Clark (ed.), *The proceedings of the 30th annual child language research forum*. Stanford: CSLI Publications.

Fillmore, C. J., Kay, P. & O'Connor, C. (1988). Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language* **64**, 501–38.

Gerken, L. (1996). Prosodic structure in young children's language production. *Language* **72**, 683–712.

Gleitman, L. & Wanner, E. (1982). Language acquisition: the state of the art. In E. Wanner & L. R. Gleitman (eds), *Language acquisition: the state of the art*. Cambridge: CUP.

Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Greenfield, P. M. & Smith, J. H. (1976). *The structure of communication in early language development*. London: Academic Press.

Hoekstra, T. & Hyams, N. (1998). Aspects of root infinitives. *Lingua* **106**, 81–112.
Hyams, N. (1992*a*). The genesis of clausal structure. In J. M. Meisel (ed.), *The acquisition of verb placement: functional categories and V2 phenomena in language acquisition*. Dordrecht: Kluwer.
Hyams, N. (1992*b*). Morphosyntactic development in Italian and its relevance to parameter-setting models: comments on the paper by Pizzuto & Caselli. *Journal of Child Language* **19**, 695–709.
Hyams, N. (1999). Underspecification and modularity in early syntax: a formalist perspective on language acquisition. In M. Darnell, K. Wheatley, F. Newmeyer, M. Noonan & E. Moravcsik (eds), *Functionalism and formalism in linguistics: volume I: general papers*. Amsterdam: John Benjamins.
Ingram, D. (1974). The acquisition of the English verbal auxiliary and copula in normal and linguistically deviant children. *ASHA Monographs* **18**, 5–14.
Kuczaj, S. A., II. (1985). General developmental patterns and individual differences in the acquisition of copula and auxiliary *be* forms. *First Language* **6**, 111–17.
Kuczaj, S. A., II. & Maratsos, M. (1983). Auxiliary verbs in yes–no questions: a different kind of category. *Developmental Psychology* **19**, 440–44.
Lieven, E. V. M., Pine, J. M. & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language* **24**, 187–219.
MacWhinney, B. (1982). Basic syntactic processes. In S. A. Kuczaj (ed.), *Language development, volume 1: syntax and semantics*. Hillsdale, NJ: Erlbaum.
MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk* (3rd edn.). Mahwah, NJ: Erlbaum.
Maratsos, M. P. & Chalkley, M. A. (1980). The internal language of children's syntax: the ontogenesis and representation of syntactic categories. In K. E. Nelson (ed.), *Children's language, volume 2*. New York: Gardner.
Peters, A. M. (1983). *The units of language acquisition*. Cambridge: CUP.
Peters, A. M. (1997). Language typology, prosody, and the acquisition of grammatical morphemes. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition, volume 5: expanding the contexts*. Mahwah, NJ: Erlbaum.
Peters, A. M. (2001). Filler syllables: what is their status in emerging grammar? *Journal of Child Language* **28**, 229–42.
Phillips, C. (1995). Syntax at age two: cross-linguistic differences. In C. T. Schütze, J. Ganger & K. Broihier (eds), *Papers on language processing and acquisition, MIT Working Papers in Linguistics* **26**, 325–82.
Pine, J. M., Lieven, E. V. M. & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics* **36**, 807–30.
Pinker, S. (1996). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Pizzuto, E. & Caselli, M. C. (1992). The acquisition of Italian morphology: implications for models of language development. *Journal of Child Language* **19**, 491–557.
Pizzuto, E. & Caselli, M. C. (1993). The acquisition of Italian morphology: a reply to Hyams. *Journal of Child Language* **20**, 707–12.
Poeppel, D. & Wexler, K. (1993). The full competence hypothesis of clause structure in early German. *Language* **69**, 1–33.
Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Oxford: Blackwell.
Radford, A. (1996). Towards a structure-building model of acquisition. In H. Clahsen (ed.), *Generative perspectives on language acquisition*. Amsterdam: John Benjamins.
Rice, M. L., Wexler, K. & Hershberger, S. (1998). Tense over time: the longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research* **41**(6), 1412–31.
Rispoli, M. (1999). Case and agreement in English language development. *Journal of Child Language* **26**, 357–72.
Rowland, C. F. & Pine, J. M. (2000). Subject–auxiliary inversion errors and wh-question acquisition: 'what children do know?' *Journal of Child Language* **27**, 157–81.

Rubino, R. & Pine, J. (1998). Subject–verb agreement in Brazilian Portuguese: what low error rates hide. *Journal of Child Language* **25**, 35–60.
Sachs, J. (1983). Talking about the there and then: the emergence of displaced reference in parent–child discourse. In K. E. Nelson (ed.), *Children's language, volume 4*. Hillsdale, NJ: Erlbaum.
Sano, T. & Hyams, N. (1994). Agreement, finiteness and the development of null arguments. *Proceedings of NELS 24*.
Santelmann, L. & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing. *Cognition* **69**, 105–34.
Schütze, C. T. (1997). *INFL in child and adult language: agreement, case and licensing*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
Schütze, C. T. & Wexler, K. (1996). Subject case licensing and English root infinitives. *Proceedings of the 20th annual Boston University conference on language development*. Somerville, MA: Cascadilla Press.
Schütze, C. T. & Wexler, K. (2000). An elicitation study of young English children's knowledge of tense: semantic and syntactic properties of optional infinitives. *Proceedings of the 24th annual Boston University conference on language development*. Somerville, MA: Cascadilla Press.
Suppes, P. (1974). The semantics of children's language. *American Psychologist* **29**, 103–14.
Tincoff, R., Santelmann, L. & Jusczyk, P. (2000). Auxiliary verb learning and 18-month-olds' acquisition of morphological relationships. *Proceedings of the 24th annual Boston University conference on language development*. Somerville, MA: Cascadilla Press.
Tomasello, M. (1992). *First verbs: a case study in early grammatical development*. Cambridge: CUP.
Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition* **74**, 209–53.
Tomasello, M. & Brooks, P. J. (1998). Early syntactic development: a construction grammar approach. In M. Barrett (ed.), *The development of language*. London: University College London Press.
Vainikka, A. (1994). Case in the development of English syntax. *Language Acquisition* **3**, 257–325.
Valian, V. (1992). Categories of first syntax: *be*, *be*+*ing*, and nothingness. In J. M. Meisel (ed.), *The acquisition of verb placement: functional categories and V2 phenomena in language acquisition*. Dordrecht: Kluwer.
Wexler, K. (1998). Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua* **106**, 23–79.