

# Predicting children's errors with negative questions: Testing a schema-combination account

BEN AMBRIDGE AND CAROLINE F. ROWLAND\*

## *Abstract*

*Positive and negative what, why and yes/no questions with the 3sg auxiliaries can and does were elicited from 50 children aged 3;3–4;3. In support of the constructivist “schema-combination” account, only children who produced a particular positive question type correctly (e.g., What does she want?) produced a characteristic “auxiliary-doubling” error (e.g., \*What does she doesn’t want?) for the corresponding negative question type. This suggests that these errors are formed by superimposing a positive question frame (e.g., What does THING PROCESS?) and an inappropriate negative frame (e.g., She doesn’t PROCESS) learned from declarative utterances. In addition, a significant correlation between input frequency and correct production was observed for 11 of the 12 lexical frames (e.g., What does THING PROCESS?), although some negative question types showed higher rates of error than one might expect based on input frequency alone. Implications for constructivist and generativist theories of question-acquisition are discussed.*

*Keywords:* language acquisition; negative questions; constructivist account; schema-combination; frames.

---

\* Our thanks are due to the children, parents and teachers at the nursery of St. Catherine’s Roman Catholic Primary School, Manchester, where the study was conducted. This research was funded by an ESRC Postdoctoral Fellowship award to Ben Ambridge (PTA-026-27-0705) and an ESRC Grant to Ben Ambridge, Julian M. Pine and Caroline F. Rowland (RES-000-22-1540). Correspondence address: Ben Ambridge, Department of Psychology, University of Liverpool, Bedford Street South, Liverpool, L69 7ZA, UK. Telephone +44 151 794 1111. Author’s e-mail: <Ben.Ambridge@Liverpool.ac.uk>.

## 1. Introduction

The key test for any theory of language acquisition is whether it can successfully predict particular patterns of error found in children's production data. Errors of commission (e.g., English past tense *-ed* or plural *-s* over-regularization errors) are particularly useful in this regard, as errors of omission may be a consequence of, for example, simple memory limitations, and not of an interesting property of the child's grammar.

Although very many theories make predictions about whether children, as a group, will display different error rates for different lexical items (e.g., regular vs irregular past tense forms, or high vs low frequency verbs) or syntactic constructions (e.g., transitive vs intransitive), few proposals are well-specified enough to make predictions regarding the particular error types that individual children will make for particular sentence types. The goal of the present study was to investigate whether a version of constructivist theory can use each child's *correct* productions for a particular sentence type (positive questions) to attribute to that child an inventory of lexically-specific frames<sup>1</sup> that can predict whether or not that child will make a particular *error* for a related sentence type (negative questions).

### 2.1. *The constructivist account*

The central idea of constructivism is that much of children's early multi-word speech consists of a network of lexically or semantically-based frames, which gradually extend in complexity over development. This idea has a long history in the field of language acquisition research. As

---

1. In this paper, we adopt the terminology of Dąbrowska and Lieven (2005). See Figure 1 for details. In particular, a *schema* is a complex symbolic unit that is made up of simpler symbolic units (complex units can be combined to form larger complex units ad infinitum, though as a simplifying assumption, we take the individual word to be the simplest unit). Each of these simple units is either *concrete* or *schematic* (also called *abstract*). A schema may consist entirely of concrete units (e.g., *open the door*), in which case it is termed a *fixed phrase*, or entirely of schematic units known as *slots* (e.g., *PROCESS THING*), in which case it is termed an *abstract schema*. A third type of schema, particularly important here, is known as a *frame* (or more properly a *lexically-specific*, or *lexically-based*, frame) and contains at least one concrete unit and at least one schematic unit (i.e., a slot). For example, in the lexically-specific frame *open THING*, *open* is the concrete unit and *THING* the schematic unit or slot. When a child uses a frame (e.g., *open THING*) to produce an actual utterance (e.g., *open the window*) we call any unit that elaborates a schematic unit (here *the window*, which elaborates the slot *THING*) a *filler*, and the process *superimposition*. For clarity, we apply these terms consistently when discussing both our own studies and those of others.

early as 1976, Braine presented a detailed analysis of 11 children's data, arguing that these children's first productive multi-word structures were confined to a small number of "formulae of limited scope for realizing specific kinds of meanings" with each formula "concerned with a specific and often rather narrow kind of semantic content" (p. 4). For example, one of the children Braine studied, Andrew, appeared to have acquired a *more+X* formula, where *more* seemed to express either observed or desired recurrence, whereas the word occupying the X position indicated the thing/event that recurred (not always grammatically; e.g., *more sock*).

Similarly, Bowerman (1976) has argued that the early utterances of her children could be explained by quite specific semantically-based schemas such as *INGEST+INGESTED* (e.g., *eat choccy*) or even more general schemas (e.g., *actor+action*). Importantly, her data did not seem to be readily interpretable in terms of abstract syntactic relations such as SUBJECT+VERB; indeed some utterances seem to rule out this interpretation. For example, Braine (1976) argued that the utterance *Mommy oops* (Bowerman 1973) could have been produced by an AGENT+ACTION but not a SUBJECT+VERB schema.

The recent constructivist approach of Lieven, Pine, Tomasello and colleagues focuses more specifically on item-based learning. On this approach, much of children's early speech consists of lexically-specific frames, with the complexity and abstraction of adult grammar coming as the endpoint of a long developmental process. For example, one of the earliest influential pieces of work in this tradition is that of Lieven, Pine and Baldwin (1997). Approximately 60% of the utterances of the 11 children they studied could be accounted for by the child's first 25 lexically-specific frames such as *There's a THING*, *I want a THING* and *ACTION it*, with many of the remaining utterances likely to be frozen phrases. This work has been extended, most notably by Lieven, Behrens, Speares and Tomasello (2003) who used a much denser longitudinal corpus of one 2-year old child to demonstrate that 63% of the utterances produced in the last hour of recording were not novel (i.e., they had been said before in exactly that form), and that 29% of the remaining utterances could be derived from one of the child's previous utterances by just one operation such as substituting a filler into a slot, or adding a word to the beginning or end of the utterance. This left merely 8% of utterances that could not be accounted for in terms of lexically-specific frames.

Most recently, Dąbrowska and Lieven (2005) have provided a thorough linguistic description of two children's early questions based on Cognitive Linguistic theory and item-based learning. Their aim was to build a linguistic grammar of two children's early utterances based on three assumptions about language: 1) that humans store symbolic units,

Table 1. Assumptions behind the item-based model of language learning (reproduced from Dąbrowska and Lieven 2005)

---

*Assumption 1: Human beings store symbolic units (i.e., pairings of a phonological form and a semantic representation).* These can be concrete and simple (*door*) or complex (*open the door*). Complex symbolic units consist of smaller units (e.g., *open, the door*), which are also pairings of a phonological form and a “chunk” of semantic structure. Symbolic units can also be partially schematic (*open NP*) or even wholly schematic (*V NP*).

*Assumption 2: Language acquisition involves the acquisition of symbolic units, both concrete and schematic.* Schematic units are generalizations over more concrete units or actual utterances. Since both kinds of units are represented in the same format and have the same structure (see also Langacker 1991, 2000), this process involves strengthening the shared features while abstracting away from the differences. For instance, the frame *Shall I PROCESS?* is a generalization over utterances such as the following:

[from Annie 3;0]

\*MOT: shall I try?

\*MOT: shall I jump in?

\*MOT: shall I look after baby?

\*MOT: shall I be Mummy?

All these utterances share certain aspects of meaning (an offer to do something) and phonological form ('[ʃælai] followed by a slot into which an expression specifying the type of activity can be inserted). Similarly, *open THING* is a generalization over expressions such as *open it, open the door, open the gate, etc.*

*Assumption 3: The production of novel expressions involves the combination of symbolic units using two operations: juxtaposition and superimposition.* JUXTAPOSITION involves linear composition of two units, one after another. Note that the two units can be combined in either order:

(3) Derivation of a novel expression using juxtaposition

*now + are you downstairs?* → *now are you downstairs?*  
or *are you downstairs now?*

*why are you holding me? + Daddy?* → *why are you holding me Daddy?*  
or *Daddy why are you holding me?*or

The linear juxtaposition signals that the meanings of the two expressions are to be integrated, but the construction itself does not spell out how this is to be done, so it must be inferred by the listener (in the first example, *now* is understood to designate the time of the situation designated by the clausal unit; in the second example, *Daddy* is the addressee).

In SUPERIMPOSITION, one unit (which we call the “filler”) elaborates a schematically specified subpart of another unit (the “frame”). For instance, the units *shall I PROCESS?* and *open that* can be superimposed to derive the novel expression *shall I open that?*.

Superimposition happens simultaneously at both the phonological and the semantic poles of the two expressions. The filler must match the properties specified in the frame: the *shall I PROCESS?* frame requires a filler which designates a PROCESS, in the technical CG sense, that is to say, a temporal relation (see Langacker 1987a, 1987b, 1991).

---

2) that language acquisition involves the acquisition of both concrete and schematic symbolic units, and 3) that the production of novel expressions involves combining symbolic units using two operations: juxtaposition and superimposition (see Table 1, reproduced from Dąbrowska and Lieven, for an explanation of these assumptions). Based on these assumptions, Dąbrowska and Lieven (2005) wrote an inventory of constructions for the question data of two children aged 2–3 years of age, using only frames and fixed phrases (i.e., no abstract schemas).

Using dense data, these authors extracted a test corpus of the child's speech (the last two transcripts recorded for one child and the last five for the other, less talkative child), and a main corpus, which contained all the remaining transcripts (both child and adult utterances). They then extracted all syntactic questions from each test corpus (the target utterances) and determined which of these target utterances were either exact repetitions of utterances in the main corpus, or could be derived from such utterances by juxtaposing or superimposing one or two component units (defined as "an expression which shares lexical material with the target and is attested at least twice in the main corpus, excluding imitations and self repeats"; p. 447). For example, the test utterance *Where can he park?* could have been derived by the superimposition of the two previously-attested complex units *Where can THING park?* and *Where can he PROCESS?* These data demonstrated that over 90% of the questions produced in the test corpus could have been derived in this way from previously-produced utterances. Importantly, the slots (e.g., PROCESS) appeared to be semantically constrained, with very few errors of the type where children superimposed a filler that was semantically inappropriate or of the wrong form-class (e.g., *\*Where can he table?*). This coincides with Braine's (1976: 9) observation that "the major source of the differences among the sets of words that occur with different pivots in positional productive patterns is obviously the meaning each particular formula is used to express... Words are not selected randomly but to accomplish a communicative purpose".

Accordingly, it is particularly important in the context of the present paper to highlight the fact that the constructivist approach investigated does *not* posit frames with unconstrained slots (e.g., *What does + X*) in which *any* word or phrase may appear, regardless of whether it is semantically appropriate<sup>2</sup>. Rather, the frames posited are of the form *What does*

---

2. The confusion arises because constructivist authors often use notation such as *X* (as opposed to *THING*) to denote a slot. This is presumably to avoid committing to either a semantic (*THING*) or a syntactic definition (*NOUN*) of a slot. In fact, constructivist

*THING PROCESS?* in order to capture the finding that children's (hypothesized) slot-fillers are virtually always semantically appropriate (e.g., *What does John want?*) and virtually never inappropriate (e.g., *\*What does John happy?*). Furthermore, the generalization procedure is constrained not only by the fact that the slot specifies (albeit in a fairly general way) the semantics of a potential filler (e.g., *happy* cannot fill the *PROCESS* slot) but also by the communicative intentions of the child. Thus a child who has the frame *What does THING PROCESS?* would be unlikely to ask a question such as *What does John want the ball?*, even though *want the ball* is potentially a valid filler for the *PROCESS* slot, simply because the utterance does not match the communicative intentions of the child (i.e., to ask the hearer what John wants).

However, these data only show compatibility with the assumptions of the item-based learning model. Dąbrowska and Lieven's (2005) analysis does not confirm that children are in fact storing these lexically-specific frames (though see Bannard and Matthews 2008) and using them to derive subsequent utterances. Thus, Rowland (2007) attempted a stronger test of this account by a detailed analysis of the pattern of correct syntactic questions and errors in 10 children's data. First, she tested the prediction that questions that could be produced by a single operational procedure (e.g., the superimposition of filler material into a frame such as *What does THING PROCESS?*) should be easier, quicker to process and thus less susceptible to error than questions requiring more complex operations or generalisation over a number of stored examples (see Rowland 2007, for rationale). Sure enough, error rates were significantly lower in questions that could have been derived from previously-learned and highly-frequent lexically-specific question frames, and this finding could not be attributed to other factors such as the identity of the *wh*-word or auxiliary, or the simple input frequency of the individual words.

---

theory does not posit that the slots are either entirely semantically or entirely syntactically defined. The slots are formed by "functionally-based distributional analysis" (Tomasello 2003); i.e., by abstracting across items that have similar distributions AND share similar communicative functions. In the present article, we use semantic labels (e.g., *PROCESS*) as opposed to syntactic labels (e.g., *VP*) to avoid committing to the notion that children are in possession of abstract syntactic categories (though, as argued above, we would not claim that the slot can be entirely semantically defined either). The one exception to this is that we use the syntactic term *AUX*, purely because no convenient "syntactically neutral" term is available (terms such as *grounding predication* are unwieldy, and likely to be unfamiliar to most researchers). This should NOT be taken as a claim that young children are actually in possession of an *AUX* category; indeed, we have argued elsewhere that they are not (e.g., Rowland and Pine 2000; Ambridge et al. 2006).

However, errors in frame-based questions, although rare, were not non-existent (also Dąbrowska and Lieven 2005 found that 20% of their potentially frame-derived questions were errors). To investigate these further, Rowland (2007) conducted a second analysis, testing whether those errors that do occur in lexical-frame-based questions (which are predicted to be rare) could be interpreted as resulting from schema-combination operations such as the superimposition of a frame (e.g., *Where does THING PROCESS?*) and an inappropriate complex concrete element (e.g., *he goes*; yielding *\*Where does he goes?*). Unfortunately, the results for this second prediction were suggestive but not decisive: In all analyses, the majority of errors in frame-based questions seemed to result from the superimposition of a frame and an inappropriate filler, but the data were too sparse for definitive conclusions to be drawn.

Thus this second prediction is still to be tested. A useful approach when naturalistic data are too sparse to allow firm conclusions to be drawn is to turn to experimental procedures where the relevant structures can be elicited in relatively large numbers. The aim of the present study, therefore, is to use an elicited-production paradigm to investigate whether particular question-formation errors can be explained by the superimposition of a lexically-specific question frame and an inappropriate filler. In order to maximize the likelihood of obtaining sufficient data to test this prediction, the present study focussed on one of the most frequent error types: auxiliary-doubling errors in negative questions (e.g., *\*What does she doesn't like?*). To provide a particularly stringent test, we decided to investigate whether the item-based-learning account can predict *which particular children* will and will not make this error, given the inventory of lexically-specific frames attributed to them on the basis of an independent set of data (positive questions).

## 1.2. *Testing the schema-combination explanation of doubling errors*

One of the most frequent errors of commission reported in the literature (e.g., Guasti et al. 1995; Rowland 2007) is the auxiliary-doubling error; in which children produce questions with two auxiliaries—the first being a preposed auxiliary and the second occurring post-subject. The majority of these are of the form *positive auxiliary (AUX) . . . negative auxiliary (AUX+NEG; e.g., \*What does she doesn't like?*). Other auxiliary-doubling errors are significantly less frequent. For example, in the study by Guasti et al., *AUX+NEG . . . AUX+NEG* errors (e.g., *\*What doesn't she doesn't like?*) constituted only 8.3% of questions and *AUX+NEG . . . AUX* errors were unattested. Guasti et al. (1995) did not report data for positive questions, but both naturalistic-data and elicited-production studies (see Ambridge

et al. 2006, for a review) have found that auxiliary-doubling errors are rare for simple positive questions (though common for multiple-clause questions; see Ambridge et al. 2008).

Under the schema-combination account AUX...AUX+NEG errors (e.g., *\*What does she doesn't like?*) are derived from lexically-specific frames (e.g., *What does THING PROCESS?*) either by mutual-elaboration with another lexically-specific frame (e.g., *she doesn't PROCESS*) or by the superimposition of a complex concrete element (e.g., *she doesn't like peas*); with each learned from (or abstracted across) declarative sentences in the input.<sup>3</sup> This means that the model makes a very strong prediction about the patterning of such errors. Specifically, such errors should occur *only* in children who already possess the relevant lexically-specific positive frame (e.g., *What does THING PROCESS?* [or the slightly more concrete frame *What does she PROCESS?*; the predictions of the account are the same in either case]).

One part of this prediction is that children who show evidence of having the relevant positive frame (e.g., *What does THING PROCESS?*) will make auxiliary-doubling errors for negative questions (e.g., *\*What does she doesn't like?*). However, because such children may also have the relevant negative frame (e.g., *What doesn't THING PROCESS?*), or indeed a more general schema that allows them to ask negative questions (e.g., *WH-WORD AUX+NEG THING PROCESS?*), a clearer prediction is that children who do NOT have the relevant positive frame (e.g., *What does THING PROCESS?*) will NOT be able to produce an auxiliary-doubling error for the corresponding negative question (e.g., *\*What does she doesn't like?*). Each child can be classified as either having or not having each particular lexically-specific positive frame (e.g., *What does THING PROCESS?*) on the basis of whether or not she is able to produce a correctly-formed positive question using this frame (e.g., *What*

---

3. Although the complex unit *she doesn't like* may well also serve as a component unit in a frame such as *she doesn't like THING*, this does not mean that it is unavailable for superimposition into a frame such as *What does THING PROCESS* (at least according to the account of superimposition given by Dąbrowska and Lieven 2005). Thus one does not have to posit some deletion process to explain why the child does not superimpose *What does THING PROCESS?* and *She doesn't like THING*, to yield *What does she doesn't like THING?*. However, one might reasonably ask how the child knows to choose *She doesn't like* as opposed to *She doesn't like THING* for superimposition with the frame *What does THING PROCESS?*. The answer is that the child understands her own communicative intention—to ask what the character does not like—and so chooses the more semantically-appropriate of the two potential fillers. Clearly *She doesn't like THING* is not a semantically appropriate filler, because it contains a slot that must be filled with information that the child does not have.



does she like?). Thus the crucial prediction tested in the present study is as follows:

AUX...AUX+NEG doubling errors for a particular negative question type (e.g., *\*What does she doesn't like?*) will be produced by a significantly greater proportion of children who did than who did not produce a correctly-formed positive question of the corresponding type (e.g., *What does she like?*). This is because production of a correctly-formed positive question constitutes evidence for knowledge of the individual lexically-specific positive frame (e.g., *What does THING PROCESS?*) that is required for the relevant doubling error.

We tested this prediction by eliciting 3sg positive and negative wh- and yes/no questions with *can* and *do* (see method section for details)<sup>4</sup>. It is important to emphasize that the claim of the present account is not simply that children will have greater difficulty forming negative than positive questions as the latter build on knowledge of the former. The prediction is that children who produce a *correctly-formed* positive question will be the same children who produce a particular *error* on a related negative question. Indeed, for positive/negative pairs for which the latter is significantly more frequent in the input than the former (e.g., *Why can THING PROCESS?* vs *Why can't THING PROCESS?*), the constructivist account predicts better performance on the negative form.

Of course, if the schema-combination account is to succeed, it must also be able to predict—on the basis of frame-inventories attributed to particular children—why they do and do not make other particular types of errors on positive and negative questions. These predictions and their rationale are given in Appendix 1. A fuller taxonomy of possible question-errors is given in Table 2 (see method section).

The present study also tests two further predictions of the schema-combination account. The prediction relates to correctly-formed questions. For all question-types (whether positive or negative), the schema-combination account predicts that rates of correctly-formed questions will be highest when the relevant frames (e.g., *What does she PROCESS?*, *What doesn't she PROCESS?*, *Does she PROCESS?*, *Doesn't she PROCESS?*)

---

4. To avoid the design of our study becoming too large it was necessary to restrict this investigation to two auxiliaries and to 3sg forms. Auxiliary DO and modal CAN were selected as the two auxiliaries generally agreed to display the highest rates of commission error (e.g., Ambridge et al. 2006; Rowland 2007; Rowland et al. 2005). Third person forms were chosen because first and second person forms are difficult to elicit with the paradigm used, and singular forms because the relevant combinations are presumably more frequent for singular than plural forms (e.g., *What does she > What do they*).

are of high frequency in the child's input. This prediction obtains because, under this account, many correctly-formed questions will be produced using partially-abstract frames of this type (with the remainder produced using wholly abstract schemas such as *WH-WORD AUX THING PROCESS?*); the higher the frequency of the relevant frame in the input, the greater the opportunity for it to be acquired. The converse of this prediction also holds: The higher the frequency of the frame in the input, the lower the predicted error rate.

Note that positive questions are therefore predicted to show higher-rates of correct production only to the extent that positive frames are more frequent than negative frames. For questions where the negative form is more frequent in the input (e.g., *Why can't THING PROCESS?* vs *Why can THING PROCESS?*) higher rates of correct-production are predicted for the negative than the positive form.

The final prediction of the schema-combination account relates to non-inversion errors (e.g., *\*What she does like?*; *\*What she doesn't like?* *\*She does like fruit?*; *\*She doesn't like fruit?*). Many errors of this type could potentially be accounted for by Dąbrowska and Lieven's (2005: 442) juxtaposition operation that "involves linear composition of two units, one after another" (e.g., the concrete simple unit *what* and the frame *she doesn't PROCESS?* to yield *What she doesn't PROCESS?*). The prediction is that these errors will be common in cases where this frame is frequent in the input, particularly when it is frequent relative to the appropriate negative question frame. For example (as we will see) *she doesn't PROCESS* is far more frequent than *What doesn't THING PROCESS?*; thus the account predicts a high rate of non-inversion errors (e.g., *\*What she doesn't PROCESS?*). Conversely, *she does PROCESS* is far less frequent than *What does THING PROCESS?*; thus the account predicts a low rate of non-inversion errors (e.g., *\*What she does PROCESS?*) and a high rate of correct-question formation.

### 1.3. *Generativist accounts*

Although the main aim of the present study is to test the predictions of the constructivist schema-combination account, it is also necessary to consider whether any rival generativist theories of question-formation could potentially explain the pattern of data observed. The theories of DeVilliers (1991), Stromswold (1990), Santelmann et al. (2002), Hattori (2003) and Guasti et al. (1995) all share the assumption that knowledge of subject-auxiliary inversion is part of children's innate UG endowment. Children make errors when they have yet to learn to apply this knowledge correctly to particular items that have language-specific properties which

may confuse or mislead children. Unlike the present proposal, all generativist theories predict higher rates of errors for negative than positive questions, because the former involve the manipulation of an additional functional projection (NEG), thus increasing complexity.

DeVilliers (1991) predicts that non-inversion errors (e.g., *Why she doesn't like the tea?*) will be more common for adjunct (e.g., *why, how*) than argument (e.g., *what, who*) wh-operators, as children are misled by the fact that some adjunct wh-operators (e.g., *How come [she doesn't like the tea]?*) do not trigger inversion. In support of this prediction, most studies of question acquisition (but not that of Ambridge et al. 2006) find higher error rates for *why* than other wh-operators (e.g., Erreich 1984; Kuczaj and Brannick 1979; Labov and Labov 1978; Rowland and Pine 2000), although results for *how* are more mixed, with some authors (e.g., Labov and Labov 1978; Rowland and Pine 2000) reporting high rates of inversion.

The theories of Stromswold (1990), Santelmann et al. (2002) and Hattori (2003) predict higher rates of commission error for questions with auxiliary DO (and copula BE) than any other auxiliary. According to these authors, DO questions are problematic because, unlike all other auxiliaries, DO is not present in the untransformed utterance (unless inserted for emphasis or negation), but is inserted to bear (or check) tense and agreement (e.g., *Minnie likes the tea* → *what DOES Minnie like?*). (Questions with copula BE are problematic because children have to learn that copula BE, as a main verb, inverts (e.g., *Minnie is happy* → *Is Minnie happy?*) whereas all other main verbs do not).

There is a certain amount of support for this prediction. It is generally recognized that rates of double-marking error are higher for DO than for other auxiliaries (e.g., Ambridge et al. 2006; Hattori 2003; Maratsos and Kuczaj 1978; Stromswold 1990), though this predominantly reflects children's performance with negated forms of DO. For yes/no questions, both Santelmann et al. (2002), using a sentence repetition paradigm, and Rowland (2007), in an analysis of naturalistic data, found significantly higher rates of commission error for DO than for the modal auxiliaries. However, for wh-questions, the experimental study of Ambridge et al. (2006) found only a non-significant trend in the predicted direction (for non-inversion errors), whilst the naturalistic data study of Labov and Labov (1978) found a higher rate of correctly formed questions for DO than CAN. Rowland (2007) found no difference between error rates for DO and the modal auxiliaries for wh-questions.

Focussing specifically on auxiliary-doubling errors (e.g., *\*What can she can eat?*), there is a long history in the generativist literature of treating such errors as reflecting the movement of the auxiliary from its original

position within IP to the head of CP ([V to] I to C movement) without subsequent deletion of the auxiliary from its original position (e.g., Hurford 1975). However simple copying-without-deletion accounts cannot explain why rates of auxiliary-doubling error should be different for different question types and, in particular, why they should be more frequent for negative than positive questions (other than by a general appeal to the notion of complexity).

A more sophisticated copying-without-deletion account that does address this phenomenon is that of Guasti et al. (1995). This theory is designed specifically to explain the patterning of children's errors with negative questions. In an elicited production study, these authors found that errors involving AUX...AUX+NEG doubling (e.g., *\*What does she doesn't like?*) were relatively frequent (39.7% of all questions produced), as were non-inversion errors (21.7%). Other auxiliary-doubling errors were significantly less frequent, with AUX+NEG...AUX+NEG errors (e.g., *\*What doesn't she doesn't like?*) constituting only 8.3% of questions and AUX+NEG...AUX errors unattested (data for positive questions are not reported).

Guasti et al.'s (1995) proposal, in simplified form, is that children believe that the negation marker (*not*, or the clitic form *n't*) may not raise out of IP (the authors take this to be a matter of parametric variation, as this is indeed the case for some language dialects, e.g., Paduan Italian). This creates a problem when the child uses the abbreviated *-n't* form of the negation marker cliticized onto the auxiliary at I (e.g., *doesn't*). In the adult grammar, both the auxiliary and the negation marker are moved to C, as per the normal process of subject-auxiliary inversion, or I to C movement (e.g., *What doesn't she like?*). In the child grammar, the auxiliary is moved but the negation marker remains at I (e.g., *What does she n't like?*). Since the *n't* marker requires a host, the child spells out the trace of the moved auxiliary, to give the most common error type (e.g., *\*What does she doesn't like?*). Non-inversion errors, the second most common error-type, occur when the child chooses to keep the negation marker within IP not by moving the auxiliary and stranding the *n't* marker, but by keeping the whole AUX+NEG form within IP (e.g., *\*What she doesn't like?*).

Guasti et al.'s (1995) theory, then, makes a simple and testable prediction. Children should make only errors which involve the negation marker remaining within IP (i.e., AUX...AUX+NEG doubling errors, and non-inversion errors). Errors which involve the marker leaving IP (e.g., AUX+NEG...AUX+NEG errors such as *\*What doesn't she doesn't like?*) should not occur, or at least should occur sufficiently rarely that they can reasonably be classed as "transition structures" (p. 236).

Errors with positive questions cannot be explained without some additional mechanism.

The theory of Guasti et al. (1995) makes a further prediction. Since errors are explained on the basis of a mis-set parameter, it follows that when children are able to produce negative questions correctly, they have correctly reset the parameter, and therefore, should be able to produce correct questions across different wh-operators (e.g., *what*, *why*) and different auxiliaries (e.g., *DO*, *CAN*). In the absence of longitudinal data, it is not possible to test directly Guasti et al.'s (1995) prediction that "all of the non-adult structures should disappear simultaneously" (p. 237). However, it follows from this prediction that there should be no stage in which errors are more frequent for certain wh-operator+auxiliary or auxiliary+subject combinations than others (e.g., *why doesn't* vs *what can't*). Of course, there is nothing to prevent Guasti et al. (1995) adding to their account additional factors that could explain such findings, provided that they refer to formal properties of the items in question (e.g., adjunct vs argument status of the wh-operator, language specific properties of *DO* etc.) that are visible to the generative grammar posited under such theories.

#### 1.4. *The present study*

To summarize, the primary goal of the present study was to assess whether a version of constructivist theory (the schema-combination account) can correctly predict which children will and will not make auxiliary-doubling errors in negative questions. We tested the schema-combination account's prediction that the patterning of children's auxiliary-doubling errors in negative questions will be predictable from the slot-and-frame patterns attributed to them on the basis of their performance with related question types (positive questions). Specifically, this account predicts that AUX...AUX+NEG doubling errors for a particular negative question type (e.g., *\*What does she doesn't like?*) will be produced by a significantly greater proportion of children who did than who did not produce a correctly-formed positive question of the corresponding type (e.g., *What does she like?*).

The study also tested two further predictions of the schema-combination account. The first is that children will make fewer errors (and produce more correctly-formed questions) for frames that are frequent in the input data of a representative corpus. The second is that non-inversion errors (e.g., *\*What she doesn't like?*) will be most common in cases where the component units that—under the account—are juxtaposed to yield such errors (e.g., *What* and *she doesn't PROCESS*) are frequent in this corpus.

A final goal was to investigate whether generativist theories of question formation that do not include a role for schema-combination can account for the pattern of correctly-formed questions and errors observed.

To this end, in the present study we elicited, from each of 50 children aged 3;3–4;3, 24 questions with different combinations of auxiliary (3sg *can*, 3sg *does*), wh-type (*what*, argument operator/*why*, adjunct operator/*yes-no*, no operator) and polarity (positive or negative).

## 2. Method

### 2.1. Participants

Participants were 50 normally-developing, monolingual English-speaking children (29 female and 21 male) aged between 3;3 and 4;3 (mean = 3;10) recruited from a primary school in Manchester, England. An additional seven children completed the warm up but were excluded for attempting to answer rather than ask questions (five children), or due to absence on the second day of the study (two children).

### 2.2. Design

The experiment employed a  $2 \times 3 \times 2$  within-subjects design with factors auxiliary (3sg modal *can*/3sg auxiliary *does*), wh-type (*what*, argument operator/*why*, adjunct operator/*yes-no*, no operator) and polarity (positive or negative question), for a total of 12 different question types. Two exemplars of each question type (with main verb *drink* or *eat* for CAN questions, and *like* or *want* for DO questions<sup>5</sup>) were elicited for a total of 24 trials. For the secondary predictions, the dependent measures were, for each child, the proportion of questions for each of the 12 question types that were (1) correctly formed questions, and (2) auxiliary-placement errors, depending on the statistical analysis in question. For the primary prediction, the dependent measure was simply the number of children who produced (or did not produce) at least one AUX...AUX+NEG doubling error (over the two trials) for the relevant negative question type.

---

5. Ideally, one would have used the same main verbs with both auxiliaries. However, because DO refers to habitual actions/states, and CAN to ability/permission, the use of the same verb with each would have resulted in some unnatural or pragmatically odd questions (e.g., *Does Mickey Mouse open a present?* Santelmann et al. 2002). Since all four verbs used were extremely common, it seems unlikely that, for example, children might have performed better with *like* and *want* than *eat* and *drink*.

### 2.3. Materials

A toy dog with an internal loudspeaker connected to a minidisc player was used to provide responses to children's questions (see Ambridge et al. 2006). A disc containing appropriate answers for each of the target questions was recorded, with the first author providing the voice for the dog. Three animal puppets (Minnie Mouse, a frog and a bear, with the latter two used in the warm up only) and eight "food and drink" toys (e.g., cake, fruit, dog food, coffee, milk, juice) were used to enact the scenarios about which questions were to be asked. A second minidisc player, connected to a Shure SM58 microphone, was used to record children's responses for later transcription.

### 2.4. Procedure

The child was first shown and asked to name the toys and animal puppets, which all were easily able to do. The experimenter then introduced the child to the 'talking dog' toy, and explained that the dog would speak only to answer questions which the child had put to him. The experimenter explained that he would help the child, by telling her what she should ask.

Four warm-up trials were used to introduce the child to the game of putting questions to the dog, in response to a prompt from the experimenter. All warm up trials used subject *who* questions with the frog or bear as subject and Minnie Mouse as object. Subject *who* questions were used as subject questions do not require subject-auxiliary inversion, and neither subject nor *who* questions formed part of the test battery. For each warm up trial, the experimenter selected the relevant characters, then, out of view of the child, pretended to perform the appropriate action and said "Oh no! Somebody kicked [or bit/pushed/dropped] Minnie Mouse. Let's ask the dog *who kicked her*. Can you ask the dog *who kicked her*? Say it after me '*who kicked her*?' and again '*who kicked her*?'. In the vast majority of cases the child produced an appropriate question (e.g., *Who kicked her?*). The experimenter then operated the minidisc player, in order to have the dog produce an appropriate response (e.g., *The frog kicked Minnie Mouse*), which the experimenter and/or the child then enacted with the toys.

The test phase was split into two sessions of 12 questions each. The first was presented immediately after the warm up, with the second presented approximately 24 hours later (or as close as possible). The order of the questions was varied in a pseudo-random manner by asking the child to select the food or drink toy about which the next question was to be asked.

For each target question, the experimenter produced four utterances (constituting the *prompt* for that trial). The first utterance (e.g., *Minnie is thirsty*) made some kind of general statement, establishing the background for the question to be asked, and did not include any material to be used in the target question. The second contained the appropriate subject (always *she*), auxiliary (except for positive *does* questions), with cliticized *n't* if appropriate, and object NP (except for *what* questions) (e.g., *The dog told me she can drink the milk*). The third and fourth utterances took the form of indirect questions beginning *I wonder . . .* and *Ask the dog . . .* respectively. These utterances contained all of the above elements, plus—for wh-questions—the wh-word (e.g., *I wonder why she can drink the milk*). To ensure equivalence of presentation, these sentences were identical for yes/no questions except that *if* replaced the wh-word (e.g., *I wonder if she can drink the milk*). The full text of all prompts can be found in Appendix 2.

When the child had attempted to produce an appropriate question, the experimenter operated the minidisc player, in order to have the dog produce an appropriate response, which was then enacted with the animal toys.

### 2.5. Coding

Children's questions were independently transcribed and coded by each author using the minidisc recordings. Agreement between the two coders was 98%, representing agreement on the coding of all but 24 of the 1,200 questions elicited. All of these cases were disagreements regarding coding, rather than what the child had actually said, and were resolved through discussion, resulting in clarification of the coding scheme.

Reponses were scored according to the taxonomy shown in Table 2. Questions which did not match any of these categorizations were coded as either (1) *Non-target*: The child asks a question of the wrong type (e.g., a *Why can't . . .* instead of a yes/no *Can't* question) or gives an unintelligible or irrelevant response or (2) *Unclassified*: The child is clearly attempting to produce a question of the appropriate type, but makes an error, or combination of errors, that does not fit the coding scheme (no child gave a null response for any trial). The distinction is important as asking the wrong question or giving an unintelligible response does not necessarily equate to problems with the specific question type. However, if the child is clearly attempting to produce a question of the appropriate type, and her response is intelligible, then an error clearly does reflect some difficulty with the particular question type. Accordingly non-target but not unclassified utterances are counted as missing data for the pur-



Table 2. *A taxonomy of some possible errors in questions, with example wh- and yes/no questions with auxiliary DO.*

				Wh- example	Yes/no example
Correct	Standard form			What does(n't) she like?	Does(n't) she like cake?
	Full form of neg.			What does she not like?	Does she not like cake?
Auxiliary-placement errors	Non-inversion errors			What she does(n't) like?	She does(n't) like cake?***
	Raising errors*			What she (not) likes?	She (not) likes cake?***
	Double marking or 'doubling' errors	Auxiliary doubling	AUX...AUX	What does she does like?	Does she does like cake?
			AUX+NEG...AUX+NEG	What doesn't she doesn't like?	Doesn't she doesn't like cake?
			AUX...AUX+NEG	What does she doesn't like?	Does she doesn't like cake?
			AUX+NEG...AUX	What doesn't she does like?	Doesn't she does like cake?
Different Auxiliaires	Tns/Agr doubling	Different Auxiliaires	What can('t) she does(n't) like?	Can('t) she does(n't) like cake?	
		AUX+, Main verb+	What does(n't) she likes?	Doesn't she likes cake?	
Other errors of commission	Incorrect Tns/Agr Case-marking errors			What do(n't) she likes?	Do(n't) she likes cake?
				What do(n't) she like?	Do(n't) she like cake?
Errors of commission	AUX comission Subject comission			What does(n't) her like?	Does(n't) her like cake?
				What she like?	She like cake?
				What does like?	Does like cake?

Note: An abbreviated version of this Table appeared in Ambridge et al. (2006).

\* Only possible for *do* questions

\*\* Acceptable in certain contexts (e.g., as echo questions)

\*\*\* Positive form, but not negative form, acceptable in certain contexts (e.g., as echo questions)

poses of statistical analysis. “Doubling” errors with different auxiliaries have unclear status under all accounts and so were scored as excluded (though, as double-marking errors, they are described as such in the taxonomy shown in Table 2). Further details of the coding scheme are given in Appendix 3.

### 3. Results

Table 3 shows the mean number of correct questions and errors produced by the children. Although the majority (1071) of the 1200 elicited utterances were valid attempts (i.e., NOT non-target), the mean number of valid attempts at a particular question type varied from 1.3 for *doesn't she* questions to 1.96 for *what can* and *what doesn't* questions (out of a possible total of 2, see Table 3 below). In order to control for the fact that the children attempted some question types more often than others, all data are subsequently presented as a proportion of valid attempts rather than a proportion of all elicited questions (i.e., excluding those coded as *non-target* but not those coded as *unclassified*).

Overall, 89% of the children's responses were a valid attempt at the elicited question and 69.9% of the children's valid attempts questions were grammatically correct. The most frequent types of error were the auxiliary-placement errors, which accounted for 15.1% of the children's valid attempts overall. Non-inversion errors (e.g., *What she can eat?*) and AUX...AUX+NEG doubling errors (e.g., *What can she can't eat?*) were the most frequent types of auxiliary-placement errors. Only 3.4% of valid attempts were unclassifiable according to our classification scheme. This pattern of errors mirrors that found in the previous literature.

Table 4 shows the mean percentage of valid attempts that were correctly formed questions and auxiliary-placement errors broken down by question type. It seems that the identity of the question type had a substantial effect on the error rate—with the proportion of correct questions ranging from 81% for *yes/no* questions with *does* and *can* to 53% for *what* questions with *doesn't*. The proportion of correct questions did not correlate with the number of valid attempts, suggesting that the differing error rates cannot be attributed simply to performance factors such as the difficulty of articulating particular question types (Spearman's  $Rho = .057$ ,  $N = 12$ ,  $p = ns$ ).

#### 3.1. Testing the generativist accounts

The first analysis tested whether overall error rates (and rates of correct-question formation) patterned as predicted by the various generativist

Table 3. Mean percentages of correctly-formed questions and errors

		Raw number of occurrences	MEAN % of valid attempts	SD		
Correct	Standard form	635	57.80	47.51		
	Full form of neg.	127	12.13	30.90		
Auxiliary- placement errors	Non-inversion errors	61	5.81	21.57		
	Raising errors	18	1.73	11.28		
	Double marking errors	Auxiliary doubling	AUX. . .AUX	2	0.17	2.94
			AUX+NEG. . .AUX+NEG	7	0.61	7.49
			AUX. . .AUX+NEG	69	6.59	23.20
			AUX+NEG. . .AUX	1	0.09	2.08
			Different Auxiliaires	(7)	Scored as <i>unclassified</i> errors	
	Tns/Agr doubling	1	0.09	2.08		
Tns/Agr on main verb	0	0	0			
Other errors of commission	Incorrect Tns/Agr	8	0.69	5.85		
	Case-marking errors	89	8.75	27.27		
Errors of commission	AUX commission	15	1.30	9.46		
	Subject commission	7	0.69	6.55		
Unclassified		31	3.38	15.92		
Non-target		129	Excluded from denominator			

Table 4. Mean percentage of correctly-formed questions and auxiliary-placement errors, and frequency of frames in the maternal data of the Manchester Corpus.

Frame	Frame Frequency	Correct Questions		Non-inversion errors		Raising errors		Auxiliary-doubling errors		All auxiliary placement errors		Valid Attempts (max = 2)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
What can THING PROCESS?	240	76	40.71	12	29.55	N/A	N/A	1	7.07	13	29.98	1.96	0.2
What can't THING PROCESS?	11	55	45.46	13	29.98	N/A	N/A	16	32.64	29	40.52	1.9	0.3
Why can THING PROCESS?	0	80	37.8	7	24.76	N/A	N/A	1	7.07	8	25.48	1.94	0.24
Why can't THING PROCESS?	48	71	42.96	9	28.01	N/A	N/A	5	20.82	14	33.56	1.94	0.24
Can she PROCESS?	26	81.25	38.07	5.21	18.56	N/A	N/A	0	0	5.21	18.56	1.8	0.49
Can't she PROCESS	1	68.42	44.15	5.26	15.55	N/A	N/A	10.53	28.85	15.79	30.99	1.14	1.14
What does THING PROCESS?	402	73	41.91	0	0	13	28.23	0	0	13	28.23	1.94	0.24
What doesn't THING PROCESS?	1	53	48.88	9	28.01	0	0	23	40.67	32	44.9	1.96	0.2
Why does THING PROCESS	32	78	39.33	0	0	4	17.02	1	7.07	5	18.21	1.94	0.24
Why doesn't THING PROCESS	7	65	45.46	6	23.99	0	0	13	31.64	19	37.65	1.88	0.33
Does she PROCESS?	123	81.25	38.07	0	0	3.13	16	0	0	3.13	16	1.72	0.54
Doesn't she PROCESS?	15	55.81	49.05	2.33	15.25	0	0	22.09	39.79	24.42	41.36	1.3	0.71
<b>Mean positive questions</b>	<b>137.17</b>	<b>78.25</b>	<b>39.32</b>	<b>4.04</b>	<b>12.15</b>	<b>6.71</b>	<b>20.42</b>	<b>0.50</b>	<b>3.54</b>	<b>7.89</b>	<b>22.74</b>	<b>1.88</b>	<b>0.33</b>
<b>Mean negative questions</b>	<b>13.83</b>	<b>61.37</b>	<b>45.99</b>	<b>7.43</b>	<b>23.47</b>	<b>0.00</b>	<b>0.00</b>	<b>14.94</b>	<b>32.40</b>	<b>22.37</b>	<b>38.16</b>	<b>1.69</b>	<b>0.43</b>
What can THING not PROCESS?	0	20	36.42										
Why can THING not PROCESS?	0	23	38.08										
Can THING not PROCESS?	2	28.95	44.47										
What does THING not PROCESS?	1	22	38.01										
Why does THING not PROCESS?	1	22	39.33										
Does THING not PROCESS?	4	30.23	46.47										

Note: For negative questions, the *Correct questions* total shown includes the figures for questions using the *full form of negation*, shown in the bottom six rows.

accounts under investigation (e.g., more errors for DO than CAN questions, for *why* than *what* questions or for negative than positive questions). The constructivist account predicts that error rates will vary on a frame-by-frame basis, according to children's proficiency with each frame.

A  $2 \times 3 \times 2$  repeated measures ANOVA<sup>6</sup> investigating the effect of auxiliary (*can/does*), question type (*what/why/y-n*) and polarity (+/-) on auxiliary-placement errors yielded significant main effects of question type,  $F(2,68) = 4.36$ ,  $p = 0.02$ , partial eta squared = 0.11, and polarity,  $F(1,34) = 11.52$ ,  $p = 0.002$ , partial eta squared = 0.25. Consistent with all accounts, errors were more frequent for negative questions ( $M = 21.0\%$ ,  $SE = 5.0$ ) than positive questions ( $M = 6.2\%$ ,  $SE = 2.2$ ). However, counter to the prediction of DeVilliers (1991), errors were actually significantly *less* common for adjunct *why* questions ( $M = 10\%$ ,  $SE = 3.9$ ) than argument *what* questions ( $M = 18.9\%$ ,  $SE = 4.3$ ;  $p = 0.005$ ), which—incidentally—also displayed a significantly lower error rate than yes-no questions ( $M = 11.4\%$ ,  $SE = 2.8$ ;  $p = 0.02$ ). We also have no evidence to support the claim (e.g., Stromswold 1990; Santelmann et al. 2002) that rates of auxiliary-placement error are higher for DO ( $M = 14.3\%$ ,  $SE = 3.5$ ) than CAN ( $M = 12.9\%$ ,  $SE = 3.2$ ).

As predicted by the constructivist account, a three way interaction of auxiliary, question type and polarity was observed,  $F(2,68) = 5.82$ ,  $p = 0.005$ , partial eta squared = 0.15, revealing that error rates differed on a frame-by-frame basis. Post hoc comparisons revealed that error rates did not differ significantly between positive questions (with the exception of *what does* vs *does yes/no* questions. For negative questions, two of the six differences reached statistical significance (with one marginal at  $p = 0.09$ ). The constructivist account predicts differences between error rates for particular questions only when they differ in ways that could affect the availability of a suitable frame, with input frequency one such factor. This input-frequency prediction is evaluated in the following section.

An analogous ANOVA with correctly-formed questions as the dependent measure yielded a similar (though inverse) pattern of results, the only difference being that the comparison between DO questions (71% correct,  $SE = 5.6$ ) and CAN questions (77%,  $SE = 5.7$ ) now reached significance,  $F(1,34) = 4.76$ ,  $p = 0.04$ , partial eta squared = 0.12. Whilst this is consistent with generativist accounts that predict an across-the-board deficit for DO questions, the finding of a three way interaction (for both

---

6. The mean values quoted in this section are different to those shown in the tables, as children who did not make a valid attempt at any one particular question type were necessarily excluded from the factorial analysis.

errors and correctly-formed questions) is not consistent with such accounts. Neither is this latter finding consistent with accounts based on the notion movement without deletion (including that of Guasti et al. 1995) which, at least in their current form, do not explain why different negative questions should attract different error rates.

### 3.2. *Input-frequency analysis*

The second analysis examined whether questions which could have been produced using frames that are of high frequency in children's input were associated with lower error rates (and higher rates of correct use) than lower frequency frames. Although we have no record of the actual input to which our participants were exposed, it is possible to obtain a reasonable estimate of the frequency of particular frames in the children's input by using a representative corpus, such as the Manchester Corpus (Theakston, Lieven, Pine, and Rowland 2001; available on CHILDES)<sup>7</sup>. This corpus was collected over one year and consists of 34 recordings for each of 12 children (aged 1;10–2;0 at the start of the study) and their caregivers (approximately 300,000 caregiver utterances in total). Table 4 shows the mean percentage of correctly-formed questions and auxiliary-placement errors (subdivided into non-inversion errors, raising errors, and auxiliary-doubling errors) for each of the twelve questions elicited, along with the frequency of the relevant frame in the maternal data from the Manchester corpus (summing across all 12 mothers).

In order to conduct this analysis, it is of course necessary to specify the the relevant frames in advance. The decision was taken that, for wh-questions, the frames should include a concrete wh-word and (negated) auxiliary with abstract slots for THING and PROCESS (e.g., *What does THING PROCESS?*; *What doesn't THING PROCESS?*) whilst, for yes/no questions, the frames would again include two concrete elements—this time the auxiliary and the subject—but only a single abstract slot, PROCESS (e.g., *Does she PROCESS?*; *Doesn't she PROCESS?*). Following Bybee (1995; Bybee and Scheibman 1999), Rowland (2007) argued that frames are formed on the basis of repeated utterances containing two elements where one element (the potential concrete element) occurs with low type frequency (and is, in the optimum case, invariant) and high token

---

7. Although it is possible that this input is completely different to that of the children in our study, we think it unlikely as these types of questions tend to occur in similar proportions across adults. In addition, any differences in input frequency between the Manchester-Corpus parents and the parents of the children in this study would add noise to the data and make it less likely that we find a significant correlation, if one exists.

frequency (e.g., *I'm ACTIONing it*), and the other (the potential abstract slot) is associated with relatively high type frequency, with each of the individual types (e.g., *kick, eat, etc...*) having relatively low token frequency. Applying this logic to questions yields the different wh-question and yes/no-question frames used in the present study (see Rowland 2007, for further details).

For the purpose of frequency counts, in order to be counted as an instantiation of a particular frame, an utterance had to include all the relevant elements (e.g., *Why can't you do it?* was counted as an instantiation of the frame *Why can't THING PROCESS* but *Why can't you?* was not).

In order to test the prediction that children will show lower rates of commission error, and higher rates of correct use, for questions that could be generated using frames that are frequent in the input, we calculated rank-order correlations<sup>8</sup> between frame frequency and the mean percentages of (1) correct questions and (2) auxiliary-placement errors. These correlations are illustrated in the scatterplot shown in Figure 1.

With all 12 question frames included, neither correlation with input frequency reached significance (correct questions: *Spearman's rho* = 0.39,  $p = 0.11$ , n.s.; auxiliary-placement errors: *Spearman's rho* = -0.46,  $p = 0.07$ , n.s.; all tests one-tailed). However, removing the outlier *Why can THING process?* had a substantial effect, with both correlations now reaching significance ( $rho = 0.67$ ,  $p = 0.01$ ,  $rho = -0.70$ ,  $p = 0.008$  for input frequency by correct questions and auxiliary-placement errors respectively). In other words, *why can* shows much lower rates of error (and higher rates of correct use) than we would expect given its absence in the input data.

In addition, inspection of Figure 1 reveals that three negative frames (*What doesn't THING PROCESS?*, *What can't THING PROCESS?* and *Doesn't she PROCESS?*) showed slightly lower rates of correct use (and higher error rates) than one would predict given their frequencies (i.e., they were the biggest outliers). This suggests that there is some additional

---

8. A correlation analysis constitutes a relatively strong test of this prediction as it collapses across frames of different types that may not be equivalent (e.g., *What does THING PROCESS?* and *Can she PROCESS?*) and because constructivist accounts do not necessarily predict a linear relationship between frequency and error rates. For example, there may be threshold effects such that once a frame has occurred a sufficiently large number of times in the input for it to be learned by the child, further presentations would not be expected to influence error rates (which, at some developmental stage, will reach zero). Conversely, there may be no difference in error rates between frames of very different frequency if both are still too rare to have been acquired at a particular stage.

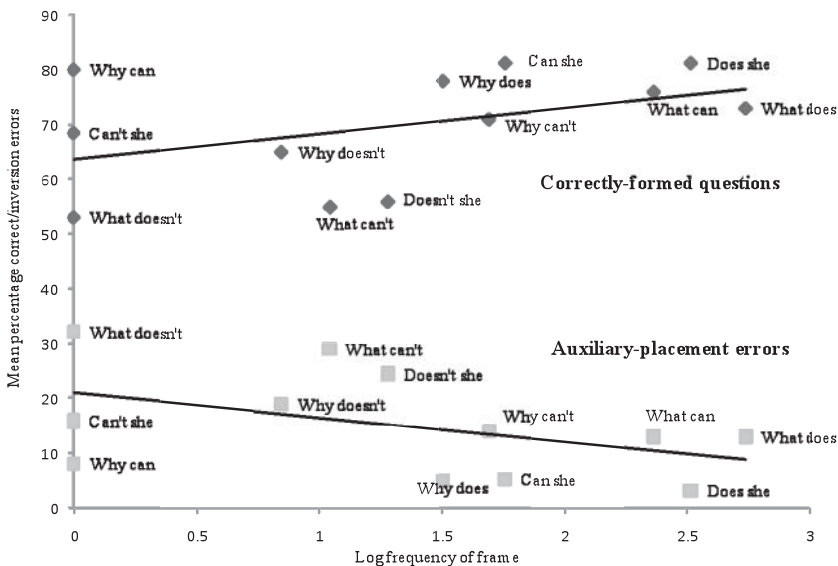


Figure 1. Relationship between (log) input frequency of the relevant lexical frame (Manchester Corpus, maternal data) and rates of (1) Correctly-formed questions and (2) Auxiliary-placement errors in the present study. Note: abstract slots (*THING PROCESS* for wh-questions and *PROCESS* for y/n questions) are not shown in the legend.

factor, beyond frequency alone, that makes negative questions particularly difficult for children, at least in this experimental setting.

It is also interesting to note that, although the absolute number of occurrences in the input is too low for a reliable statistical analysis to be performed, there is an almost perfect rank-order relationship between the input frequency of individual full-form-of-negation frames (e.g., *Does THING not PROCESS?*) and children's production of these type of questions (see Table 4). Of course, without statistical analysis, this conclusion remains highly speculative and there is likely to be considerable variation between children in the extent to which these forms are encountered in the parental input.

### 3.3. Schema-combination analysis 1: Non-inversion errors

Under the schema-combination account, an important source of non-inversion errors (e.g., *\*What she doesn't like?*; *\*Why she doesn't like the peas?*) is the juxtaposition of a one-word concrete simple unit (e.g., *what* or *why*) and a frame learned from declarative utterances (e.g., *she doesn't*



PROCESS). Thus question-types for which the relevant complex frame (e.g., *she doesn't PROCESS*) is frequent in the input should be relatively susceptible to non-inversion errors (e.g., *\*What she doesn't like?*). On the other hand, as we have already seen, if the relevant question frame (e.g., *What doesn't THING PROCESS?*) is also highly frequent in the input, then the question-type will be protected from error.

To capture this trade-off effect, we classified (in the spirit of Marchman et al. 1999; for past-tense forms) each of the 12 question-type in the present study as either at HIGH risk ( $N = 6$ ) or LOW risk ( $N = 6$ ) of non-inversion error. This was done by calculating, for each question type, the ratio of the frequency of the frame found in declaratives (e.g., *she doesn't PROCESS*) to the frequency of the relevant question frame (e.g., *what doesn't THING PROCESS*). The six question-types for which this ratio was highest were classified as being at HIGH risk of error, with the remainder (for which, for three question-types, the question frame was *more* frequent than the complex declarative unit) classified as LOW risk. The relevant values are shown in Table 5 (with all frequencies again from the maternal data of the Manchester Corpus).

This analysis offers a principled constructivist explanation of why negative questions attract particularly high rates of non-inversion error. As Table 5 reveals, all but one of the negative question-types (*Why can't THING PROCESS?*) were classified as being at high-risk of non-inversion errors, as the declarative frame that can be superimposed with a simple concrete unit (e.g., *why*) to form a non-inversion error is frequent *relative to the question frame* in the input (though questions of the form *Why can't THING PROCESS?* still retain a higher rate of non-inversion errors than one would predict on the basis of this analysis). A within-subjects *t*-test with the independent variable of risk-level (HIGH vs LOW) and the dependent variable of mean proportion of non-inversion errors (averaging across the six question types for each risk-level, for each child) revealed a significant effect of risk-level,  $t(df = 49) = 1.87$ ,  $p = 0.03$  (one-tailed test). Question-types classified as being at HIGH risk of non-inversion errors were, as predicted, associated with a higher rate of such errors ( $M = 7.60\%$ ,  $SD = 18.00$ ) than types classified as being at LOW risk of such errors ( $M = 4.67$ ,  $SD = 11.31$ ).

*Schema-combination analysis 2: Auxiliary-doubling errors in negative questions* This analysis tested the main prediction of the schema-combination account outlined in the introduction. Under this account, AUX... AUX+NEG doubling errors (e.g., *\*What does she doesn't like?*) are caused by the child superimposing a positive question frame (e.g., *What does THING PROCESS?*) and a negative declarative frame that partially

Table 5. *Classification of question-types elicited in the present study as being at either HIGH or LOW risk of non-inversion error*

Question Frame	Frequency	Declarative Frame	Frequency	Declarative/Question	Non-inversion risk
What can THING PROCESS?	240	she can PROCESS	130	0.54	LO
What can't THING PROCESS?	11	she can't PROCESS	61	5.55	HI
Why can THING PROCESS?	0	she can PROCESS	130	8	HI
Why can't THING PROCESS?	48	she can't PROCESS	61	1.27	LO
Can she PROCESS?	26	she can PROCESS	130	5.00	LO
Can't she PROCESS	1	she can't PROCESS	61	61.00	HI
What does THING PROCESS?	402	she does PROCESS	37	0.09	LO
What doesn't THING PROCESS?	1	she doesn't PROCESS	131	131.00	HI
Why does THING PROCESS	32	she does PROCESS	37	1.16	LO
Why doesn't THING PROCESS	7	she doesn't PROCESS	131	18.71	HI
Does she PROCESS?	123	she does PROCESS	37	0.30	LO
Doesn't she PROCESS?	15	she doesn't PROCESS	131	8.73	HI

Table 6. *Testing the schema-combination account: Main analysis*

Number of children producing at least one correct positive question showing evidence of frame...		Number of children producing at least one AUX...AUX+NEG doubling error showing evidence of schema-combination, yielding frame...		Fisher's Phi	<i>p</i> =
		<i>What can she can't PROCESS?</i>			
		Yes	No		
<i>What can THING PROCESS?</i>	Yes	7	13	-0.28	0.12 n.s.
	No	1	10		
		<i>Why can she can't PROCESS?</i>			
		Yes	No		
<i>Why can THING PROCESS?</i>	Yes	1	15	-0.15	0.67 n.s.
	No	0	8		
		<i>Can she can't PROCESS?</i>			
		Yes	No		
<i>Can she PROCESS?</i>	Yes	4	15	-0.21	0.44 n.s.
	No	0	4		
		<i>What does she doesn't PROCESS?</i>			
		Yes	No		
<i>What does THING PROCESS?</i>	Yes	11	12	-0.38	<b>0.03 *</b>
	No	1	10		
		<i>Why does she doesn't PROCESS?</i>			
		Yes	No		
<i>Why does THING PROCESS</i>	Yes	7	11	-0.42	<b>0.04 *</b>
	No	0	9		
		<i>Does she doesn't PROCESS?</i>			
		Yes	No		
<i>Does she PROCESS?</i>	Yes	9	13	-0.39	<b>0.03 *</b>
	No	0	8		

elaborates the slots of this first frame (e.g., *she doesn't PROCESS*). The account therefore predicts that AUX...AUX+NEG doubling errors for each particular negative question type (e.g., *\*What does she doesn't like?*) should be produced by a significantly greater proportion of children who do than do not show evidence (from their positive-question data) of having acquired the relevant positive-question frame (e.g., *What does THING PROCESS?*). Table 6 shows, for each positive/negative question pair, the number of children who (by columns, then rows)

- (Cell 1) Produced at least one correctly-formed positive question AND at least one AUX...AUX+NEG doubling error for the negative question. *Predicted to occur.*

- (Cell 2) DID NOT produce a correctly-formed positive question BUT DID produce at least one AUX...AUX+NEG doubling error for the negative question. *Predicted not to occur.*
- (Cell 3) Produced at least one correctly-formed positive question BUT DID NOT produce an AUX...AUX+NEG doubling error for the negative question.
- (Cell 4) Produced neither a correctly-formed positive question NOR an AUX...AUX+NEG error for the negative question.

The schema-combination account makes no prediction regarding correctly-formed negative questions as, under the account, these are formed using a negative question schema (either entirely schematic or partially concrete) and not by combining schemas. Therefore, only children who produced at least one error (of any type) for the appropriate negative question were included in this analysis. Children who did not produce a valid attempt at both the positive and negative question of a pair (e.g., at both a *what does* and a *what doesn't* question) were also excluded. Finally, children who produced exclusively full-form of negation responses (e.g., *What does she not like/want?*) for a particular negative question type were also excluded, as the status of such utterances is unclear (they may have been formed creatively by combining a positive frame and the negation marker *not* or may simply utilize a frame learned from the input). Applying these criteria meant that, for any one positive/negative pair, around half the children were excluded from the analysis.

The first analysis investigated whether significantly more of the six question pairs than one would expect by chance (i.e., significantly  $> 3/6$ ) followed the predicted pattern of AUX...AUX+NEG doubling errors being produced by a greater proportion of children who do than do not show evidence of having acquired the relevant positive question frame. All six positive/negative pairs followed the predicted pattern, significantly different to chance (i.e.,  $3/6$ ) by sign test,  $p < 0.02$ . Indeed, of the 69 AUX...AUX+NEG errors produced in the present study only two (each made once each by a single [different] child) were made by children who did not show evidence of having acquired the relative positive question frame. This proportion (i.e.,  $2/69$ ) is significantly lower than the proportion that one would expect by chance if there were no relationship between having acquired the relevant positive frame and producing an AUX...AUX+NEG doubling error (i.e.,  $34.5/69$ ),  $p < 0.000001$  by binomial test.

The strong prediction of the account is that, for each particular positive/negative pair individually, AUX...AUX+NEG doubling errors for the negative question should be produced by a significantly greater

proportion of children who do than do not show evidence of having acquired the relevant positive-question frame. Although the proportions were in this direction for all six pairs, the likelihood of finding a significant effect for each pair individually is reduced by the fact that we were able to include only between 23 and 34 of the total sample of 50 participants in each analysis. In fact, no analysis included a sufficient number of participants for a chi-square analysis to be legal (i.e., expected frequencies were  $<5$  for some cells). Table 6 therefore shows the results of Fisher's exact tests conducted for each positive/negative question pair.

Despite the small number of participants in each cell, for three of the six pairs (all those involving *does*) this analysis reached statistical significance at  $p < 0.05$ . For the remaining three pairs, although the analysis did not reach statistical significance,<sup>9</sup> it is important to remember that—across all three pairs—only a single utterance (*\*What can she can't eat?*) violated the prediction of the account (i.e., was produced by a child who did not produce a correct question of the form *What can THING PROCESS?*).<sup>10</sup>

---

9. It is tempting to read much into the fact that the effect is much bigger (and significant across the board) for questions with *doesn't* than questions with *can't*. However, this difference is almost certainly due to the fact that the children produce many more doubling errors with *doesn't* than with *can't*. In other words, the lack of effect for questions with *can't* is most likely due to a paucity of data.

10. A reviewer suggested that one would ideally want to show that knowledge of the relevant positive frame is a better predictor of whether or not children will make an AUX...AUX+NEG error for a particular negative question type than is knowledge of other positive frames or rates of AUX...AUX+NEG error for other negative question-types. In principle, this could be done using a binary logistic regression analysis with the outcome variable of whether or not each child made an AUX...AUX+NEG error for each question type, and predictor variables of (1) whether or not the child produced a correctly-formed positive question of the relevant type (a categorical predictor), (2) the total number of correctly-formed positive questions and (3) the total number of AUX...AUX+NEG errors (in both cases for the remaining question-types only). In practice, however, a regression analysis cannot be performed when there is a high correlation between the different predictor variables, as there is likely to be in this case: Children who have learned any one particular positive frame are likely to have learned more different positive frames than children who have not learned this frame. Similarly, children who adopt a schema-combination strategy for one negative question type are likely to use this strategy for other negative question types. An additional problem is that a very large number of participants would be needed in order to have a sufficient number in each cell of the chi-square. In any case, even without conducting the analysis discussed here, it is difficult to see how an artefactual explanation of the present findings based on individual differences in overall rates of correct-question production and/or AUX...AUX+NEG errors could give rise to the precise pattern of data observed.

#### 4. Discussion

The present study used an experimental paradigm to elicit 3sg positive and negative *what*, *why*, and *yes/no* questions with auxiliaries CAN and DO from children aged 3;3–4;3. The study demonstrated that different question types attracted different rates and different types of error, a fact that could be explained in terms of input frequency for many, but not all, of the question types investigated here. The particular patterning of errors was consistent with generativist accounts that suggest particular difficulties with auxiliary DO (e.g., Stromswold 1990; Santelmann et al. 2002), though not those that suggest that adjunct wh-words (e.g., *why*) are problematic (e.g., De Villiers 1991). The generativist theory of Guasti et al. (1995) also received mixed support. As predicted, AUX... AUG+NEG errors (e.g., *What does she doesn't like?*) were the only doubling errors that occurred with any frequency. On the other hand, this theory in its current form cannot explain the finding of different error rates for different question types, less still why this should show a moderate correlation with the frequency of the relevant question frame in the input.

The constructivist schema-combination account fared somewhat better, though support was still mixed. In support of the account, the frequency of particular frames in the input showed a modest correlation with rates of correct production and auxiliary-placement errors. Furthermore, the relative frequencies of the question frame (e.g., *What doesn't THING PROCESS?*) and a declarative frame that could yield non-inversion errors (e.g., *she doesn't PROCESS*) could be used to predict, to some extent, the likelihood of such errors. On the other hand, negative questions in general did seem to be associated with higher error rates than one would expect, given either the absolute or relative frequencies of the relevant frames.

In support of this account, AUX... AUX+NEG doubling errors for a particular negative question type (e.g., *\*What does she doesn't like?*) were produced by a significantly greater proportion of children who did than who did not produce a correctly-formed positive question of the corresponding type (e.g., *What does THING PROCESS?*). This provides support for the accounts' claim that such errors result from the superimposition of a positive question frame (e.g., *What does THING PROCESS?*) and a negative declarative frame (e.g., *she doesn't PROCESS*).

Together, these findings provide some support for the particular constructivist account under investigation here and, more generally, for accounts that include a role for lexical content. In fact, it is difficult to see how an account that includes no role for lexical content could account for

this fine-grained pattern of results. The pattern cannot be explained by a factor such as general competence, as the finding is that children who produced AUX...AUX+NEG doubling ERRORS are generally those who CORRECTLY produced the corresponding positive question (only two children produced an AUX...AUX+NEG doubling error without producing a corresponding correctly-formed positive question).

It is possible (indeed, likely) that production of a correctly formed positive question and an AUX...AUX+NEG error are both correlated with a factor such as working memory. It is also likely that correctly-formed positive questions and corresponding AUX...AUX+NEG doubling errors prime one another, as they share considerable lexical material (see Savage et al. 2003). It could be, therefore, that it is overstating the case to say that AUX...AUX+NEG doubling errors are caused by children combining schemas that they have stored in memory. Perhaps these errors have another source, but are made more likely by the recent production of a correctly-formed positive question with the relevant lexical frame. Nevertheless, it remains the case that accounts that include no role for lexical chunks are unlikely to be able to explain the pattern of data observed in the present study.

Indeed, the finding of lexical effects in the patterning of children's errors in question production is by now relatively uncontroversial. The elicited production study of Ambridge et al. (2006) found a role for specific wh-operator+auxiliary frames. More generally, lexical effects in question acquisition can be found even in reports of experimental studies whose authors perhaps favour a more formalist approach (e.g., Erreich 1984; Santelmann et al. 2002; Valian and Casey 2003; see Ambridge et al. 2006: 554–556 for a review). Looking at naturalistic data studies, effects of lexical frames are reported by Fletcher (1985), Rowland and Pine (2000) and Rowland (2007). Again, more generally, different error rates for different wh-operators and auxiliaries are reported by authors from a range of different theoretical perspectives (e.g., Tyack and Ingram 1977; Labov and Labov 1978; Maratsos and Kuczaj 1978; Stromswold 1990). Thus, studies on both naturalistic and experimental data suggest a role for lexical context in the formation of correct questions and errors.

However, there is one important discrepancy between the findings of our experimental study and those of naturalistic-data studies. As Table 2 above shows, for both auxiliary and polarity types, the children in this study produced more auxiliary-placement errors with *what* questions than with *why* questions. This stands in contrast to the findings of many naturalistic studies and to the consensus in the literature that *why* attracts much higher rates of error than other wh-words (e.g., Erreich

1984; Kuczaj and Brannick 1979; Labov and Labov 1978; Rowland and Pine 2000). For example, Rowland and Pine (2000) report a rate of subject-auxiliary non-inversion errors of 91.7% for *why* questions, as compared with only 21.4% for *what* questions. This difference in the data from naturalistic and elicited speech requires explication.

In naturalistic data, the majority of *why* questions are negative questions, whereas virtually all *what* questions are positive (Rowland and Pine 2000). Since, as the present study shows, children generally have far more difficulty with negative than positive questions, this confound may go some way to explaining the apparent effect in naturalistic data. However, this cannot account for the discrepancy between the present study and naturalistic findings, as the higher rate of non-inversion errors for *what* than *why* questions in the present study holds across both positive and negative questions.

The answer may lie in the specific questions that children produce. In naturalistic data, by far the most frequent *what* question is *What's that?* (see Ambridge et al. 2006: 546–547). When questions are counted by tokens not types, repeated production of this (probably rote-learned) phrase will drive down the rate of non-inversion error. This point was illustrated by Labov and Labov (1978) who showed that rates of non-inversion error with *what* questions climbed from 23% to 80% when questions with contracted auxiliaries were excluded. Thus, the results of the present study do not directly contradict those from analyses of naturalistic data. Rather, they highlight the importance of controlling for potentially rote-learned forms when calculating error rates (see Rubino and Pine 1998).

Finally, two findings of the present study are not well explained by the schema-combination account; the low rate of error with *why can* questions (despite its absence in the input sample) and the high rate of error with certain types of negative questions. In fact, constructivist approaches in general have difficulty in explaining why children show particularly high error rates for negative questions. Although it is true that negative questions are generally less frequent than their positive counterparts, we have seen that frequency is by no means a perfect predictor of error rate. Indeed, for the only question pair for which the negative version is more frequent in the input than the positive version (*Why can't* vs *Why can*; 49 vs 0 occurrences), the error rate is still higher for the negative question. Thus the findings of the present study suggest that there is probably a more general problem with negation than can be explained by frequency alone. Similarly, the children were able to produce *why can* questions correctly, despite such questions being of very low frequency in their input. This finding suggests an ability to generalise from other, correctly produced, high frequency positive questions types.



Although constructivist theories do not claim that frequency is the ONLY determinant of error rates (e.g., Rowland and Pine 2003), they are currently underspecified with regard to other factors (such as the presence or absence of negation) that seem to affect error rates. That said, the finding of the present study that the relative frequencies of the question frame (e.g., *What doesn't THING PROCESS?*) and a declarative frame that could yield non-inversion errors (e.g., *she doesn't PROCESS*) can, to some extent, predict the likelihood of such errors (e.g., *\*What she doesn't PROCESS?*) illustrates a more sophisticated frequency-based explanation of the type that constructivist accounts should pursue.

In conclusion, the findings of this study provide support for—but cannot be entirely explained by—the schema-combination account. It seems likely that errors are the result of superimposition and juxtaposition operations conducted over stored schemas, but it is also clear that differences in input frequency cannot entirely explain the different availability and usage of different schemas. It is clear, then, that future constructivist accounts of question-acquisition will need to move beyond input frequency, and seriously consider other factors that influence error rates, such as the presence or absence of negation. It is equally clear that the existence of lexical effects in the data is now almost undeniable, and thus that future generativist accounts will have to incorporate some role for the learning of particular lexical strings (e.g., *What does*), if not constructivist-style frames (e.g., *What does THING PROCESS?*).

Received 15 October 2007

University of Liverpool

Revision received 5 September 2008

## Appendix 1: Further predictions of the schema-combination account

Form: AUX+NEG...AUX+NEG Error

Example: *\*What doesn't she doesn't like?; \*Does she doesn't like the fruit?*

Prediction: Error very infrequent for all children

Rationale: Children who are in possession of a negative schema (e.g., *What doesn't THING PROCESS?*) will not make the error because knowledge of this frame includes knowledge that the *PROCESS* slot is filled with a positive polarity item in simple *non-finite* form\* (e.g., *like*). Children who are not in possession of a negative schema (e.g., *What doesn't THING*

*PROCESS*) will be unable to produce any form of the relevant negative question that begins with this material (e.g., *What doesn't*). Children will not produce this error as an attempt at a positive question because they understand the semantics of the negation operator *n't* (or *not*) and so will not entertain the use of frames containing this unit (e.g., *What doesn't THING PROCESS?*) when attempting to produce a positive question.

Form: AUX+NEG...AUX Error  
 Example: \**What doesn't she does like?*; \**Doesn't she does like the fruit?*  
 Prediction: Error very infrequent in all children  
 Rationale: Same as for AUX+NEG...AUX+NEG errors.

Form: AUX...AUX Error  
 Example: \**What does she does like?*; \**Does she does like the fruit?*  
 Prediction: Error very infrequent in all children  
 Rationale: Children who are in possession of a positive schema (e.g., *What does THING PROCESS?*) will not make the error because knowledge of this frame includes knowledge that the PROCESS slot is filled with a (positive polarity) item in simple non-finite form<sup>a</sup> (e.g., *like*). Children who are not in possession of a positive schema (e.g., *What does THING PROCESS?*) will be unable to produce any form of the relevant positive question that begins with this material (e.g., *What does*)

Form: Correct full-form negative question  
 Example: *What does she not like?*; *Does she not like the fruit?*  
 Prediction: Unclear, as this is an acceptable dialectal form in the North-West of England (though presumably not in all dialects, as Guasti et al. 1995, classify such forms as errors).  
 Rationale: Assuming that children are learning this form from their input, the predictions/rational are the same as for correctly-formed positive and negative questions (i.e., children form frames such as *What does THING not PROCESS?*). A reviewer suggested that children might form a more abstract schema such as *WH-WORD AUX THING NEG PROCESS* that could account for both these forms (e.g., *What does she not like?*) and for AUX...AUX+NEG errors (e.g., *What does she doesn't like?*); this latter case by additionally assuming that children make a lexical error, using *don't* etc. in the

place of *not*. This is possible in principle, but the fact that *don't* and *not* are likely to be learned in different frames makes this lexical error improbable in our view (e.g., one would not expect a child to say \**That's don't fair* instead of *That's not fair*). In any case, it is difficult to address these issues without knowing which children were exposed to negative questions of this type (though dialectal variations in negative questions may be an interesting future study).

Form: Raising error<sup>b</sup>

Example: \**What she likes?*, \**What she not likes?*, \**She likes cake?*, \**She not likes cake?*

Prediction: Same as for non-inversion errors

Rationale: A similar process is predicted as for non-inversion errors (see main text). For example a child might combine the simple concrete unit *what* and the frame *she PROCESS* (or the complex concrete unit *she likes*) to give \**What she likes?* Again, these errors will be most common when the second unit (e.g., *she likes*) is frequent, particularly when it is frequent relative to the appropriate question frame (e.g., *What does THING PROCESS?*)

Other possible errors include tense-doubling (e.g., \**What doesn't she likes?*), agreement errors (e.g., \**What do she likes?*), auxiliary-omission (\**What she like?*) and subject-omission (\**What does like?*). All of these errors could, in principle, be generated by schema-combination operations but were produced too infrequently in the present study to allow predictions to be tested. Case-marking errors (e.g., \**What does her like?*), which were relatively common, could in principle be generated by schema combination, but most likely have a different source (see Ambridge and Pine 2006)

<sup>a</sup>By this we do not mean that children have full command of a generativist-style system of TENSE and AGREEMENT, simply that they have formed a semantically/distributionally-based category of items that appear in this position (e.g., *like, want, eat, drink* etc. . .).

<sup>b</sup>“Raising” errors are those where tense is marked on the main verb and not on the dummy auxiliary *do*. We do not mean to imply that such errors actually reflect failure to “raise” tense from the main verb inside VP (as under a generativist analysis), but no suitable alternative term suggests itself.

**Appendix 2: Full text of all prompt sentences used during the study**

VERB	Wh	Pol	No	Statement	I wonder...	Ask the dog...
Can	What	+	1	Minnie is hungry. The dog told me there's some food here <b>she can eat.</b>	I wonder <b>what she can eat</b>	Ask the dog <b>what she can eat</b>
		-	2	Minnie isn't allowed all this food. The dog told me to have the food <b>she can't eat.</b>	I wonder <b>what she can't eat</b>	Ask the dog <b>what she can't eat</b>
	Why	+	3	Minnie is hungry. The dog told me <b>she can eat the cheese.</b>	I wonder <b>why she can eat the cheese</b>	Ask the dog <b>why she can eat the cheese</b>
		-	4	Minnie isn't allowed all this food. The dog told me <b>she can't eat the biscuits.</b>	I wonder <b>why she can't eat the biscuits</b>	Ask the dog <b>why she can't eat the biscuits</b>
	Y-N	+	5	Minnie is hungry. I think the dog told me <b>she can eat the crisps.</b>	I wonder <b>if she can eat the crisps</b>	Ask the dog <b>if she can eat the crisps</b>
		-	6	Minnie isn't allowed all this food. I think the dog told me <b>she can't eat the banana.</b>	I wonder <b>if she can't eat the banana</b>	Ask the dog <b>if she can't eat the banana</b>
Does	What	+	7	Minnie is hungry. The dog told me there's some food here <b>she likes.</b>	I wonder <b>what she likes</b>	Ask the dog <b>what she likes</b>
		-	8	Minnie isn't allowed all this food. The dog told me to have the food <b>she doesn't like.</b>	I wonder <b>what she doesn't like</b>	Ask the dog <b>what she doesn't like</b>
	Why	+	9	Minnie is hungry. The dog told me <b>she likes the cake.</b>	I wonder <b>why she likes the cake</b>	Ask the dog <b>why she likes the cake</b>
		-	10	Minnie isn't allowed all this food. The dog told me <b>she doesn't like the vegetables.</b>	I wonder <b>why she doesn't like the vegetables</b>	Ask the dog <b>why she doesn't like the vegetables</b>
	Y-N	+	11	Minnie is hungry. I think the dog told me <b>she likes the biscuits.</b>	I wonder <b>if she likes the biscuits</b>	Ask the dog <b>if she likes the biscuits</b>
		-	12	Minnie isn't allowed all this food. I think the dog told me <b>she doesn't like the fruit.</b>	I wonder <b>if she doesn't like the fruit</b>	Ask the dog <b>if she doesn't like the fruit</b>

VERB	Wh	Pol	No	Statement	I wonder...	Ask the dog...	
Can	What	+	13	Minnie is thirsty. The dog told me there's something here <b>she can drink</b> .	I wonder <b>what she can drink</b>	Ask the dog <b>what she can drink</b>	
		-	14	Minnie isn't allowed all of these drinks. The dog told me to have the ones <b>she can't drink</b> .	I wonder <b>what she can't drink</b>	Ask the dog <b>what she can't drink</b>	
	Why	+	15	Minnie is thirsty. The dog told me <b>she can drink the milk</b> .	I wonder <b>why she can drink the milk</b>	Ask the dog <b>why she can drink the milk</b>	
		-	16	Minnie isn't allowed all of these drinks. The dog told me she <b>can't drink the lemonade</b> .	I wonder <b>why she can't drink the lemonade</b>	Ask the dog <b>why she can't drink the lemonade</b>	
	Y-N	+	17	Minnie is thirsty. I think the dog told me <b>she can drink the water</b> .	I wonder <b>if she can drink the water</b>	Ask the dog <b>if she can drink the water</b>	
		-	18	Minnie isn't allowed all of these drinks. I think the dog told me <b>she can't drink the tea</b> .	I wonder <b>if she can't drink the tea</b>	Ask the dog <b>if she can't drink the tea</b>	
	Does	What	+	19	Minnie is thirsty. The dog told me there's something here <b>she wants</b> .	I wonder <b>what she wants</b>	Ask the dog <b>what she wants</b>
			-	20	Minnie isn't allowed all of these drinks. The dog told me to have the drinks <b>she doesn't want</b> .	I wonder <b>what she doesn't want</b>	Ask the dog <b>what she doesn't want</b>
Why		+	21	Minnie is thirsty. The dog told me <b>she wants the coke</b> .	I wonder <b>why she wants the coke</b>	Ask the dog <b>why she wants the coke</b>	
		-	22	Minnie isn't allowed all of these drinks. The dog told me <b>she doesn't want the orange</b> .	I wonder <b>why she doesn't want the orange</b>	Ask the dog <b>why she doesn't want the orange</b>	
Y-N		+	23	Minnie is thirsty. I think the dog told me <b>she wants the coffee</b> .	I wonder <b>if she wants the coffee</b>	Ask the dog <b>if she wants the coffee</b>	
		-	24	Minnie isn't allowed all of these drinks. I think the dog told me <b>she doesn't want the juice</b> .	I wonder <b>if she doesn't want the juice</b>	Ask the dog <b>if she doesn't want the juice</b>	

**Appendix 3: Further details of the coding scheme***Substitutions/omissions*

Substitutions of verbs that did not change the structure of the question (e.g., *eat* for *drink*, *want* for *like*) were scored as if no such substitution had occurred. Irrelevant errors such as determiner omission were also ignored. Substitutions of auxiliaries were not allowed, and were scored as *non-target* questions. For six trials, children substituted *he* for *she*. Since all of these trials contained additional errors (one of verb agreement, one of subject omission, one of non-inversion, and three non-classifiable errors) the decision was taken to ignore the substitution, since the utterances were already scored as erroneous (and therefore did not form the basis for crediting children with a particular frame).

For 27 questions, children used a full NP (Minnie or Minnie Mouse) instead of the target *she*. Whilst such substitutions were, of course, permitted, they potentially present a problem for testing the predictions the schema-combination account with regard to *yes/no* questions, as children are claimed to be using frames such as *Can she PROCESS?* (we assume children did not have pre-learned schemas such as *Can Minnie Mouse PROCESS?*). Fortunately, none of the children who used a full subject NP produced a correctly formed question with this NP (perhaps providing some support for the schema-combination account). One of these 27 questions exhibited an auxiliary-doubling error (*Can Minnie Mouse can't eat the bread?* Child 20, Question 6), and was therefore excluded from the schema-combination analysis only.

*Ambiguous auxiliary-doubling errors*

Seven “auxiliary-doubling” errors in fact used two different auxiliaries (e.g., *Does she can't drink the tea?*; *Is Minnie Mouse doesn't want the juice?*). Since, in every case, the second auxiliary and not the first was the target auxiliary (and so did not provide evidence of the child having formed a positive schema) these utterances were coded as *Unclassified* (not *Non-target*, as the fact that the second auxiliary was the intended auxiliary demonstrates that children were attempting to produce a question of the appropriate type).

*Applying mutually-exclusive categorization criteria*

Scoring was complicated by the fact that virtually all questions that contained an error of case marking (*her* for *she*) or verb agreement (*do[n't]* for *does[n't]*) also contained another error, usually a non-inversion/raising error, but often an auxiliary-omission error and, in one case, an auxiliary-doubling error. In order for our statistical analyses to be valid,

however, it is necessary that the classification categories are mutually exclusive. This was achieved by applying categorization in the following order:

1. Non-target/Unclassified
2. Verb agreement error
3. Case marking error
4. All other errors

For example, the questions *Why her don't want an orange?* (for *Why doesn't she...*) and *Her don't like the fish?* (for *Doesn't she...*), which appear to contain errors of verb agreement, case marking and non-inversion were classified as errors of verb agreement, whilst the question *Why her can't eat the biscuits*, which displays errors of case marking and non-inversion, was classified as a case marking error. Trials classified as non-target/unclassified were coded as such, with no further coding.

Whilst this results in losing some fine-grained detail, it does not affect the predictions of any of the theories under test. From a generativist viewpoint, failure to correctly mark subject or verb agreement is evidence of a movement failure (the child has failed to raise the item to AGRs). Thus whether a response is classified as an non-inversion/raising error or an error of verb or noun agreement marking makes little difference, as the generativist theories make predictions about the particular constructions (*why* vs *what*, DO vs CAN) for which movement (whether to C or to AGRs) will prove difficult. From a constructivist viewpoint, errors of both verb agreement/case marking and non-inversion/raising reflect the child's lack of an appropriate frame. The ordered application of these coding criteria resulted in one AUX...AUX+NEG error that could have resulted from the application of a positive frame (*Why does she don't like the peas?*, Child 13, Question 10) being coded as a verb agreement error and not as a doubling error, thus slightly reducing the likelihood of observing our predicted effect.

#### *Non-target questions*

A difficult issue was that of *non-target* questions, produced for eighty (10%) of the 816 individual trials. The most common scenario was that where the child asked a related wh-question instead of a yes/no question (e.g., *Why can't she drink the the tea* for *Can't she drink the the tea?*), perhaps because wh-questions were twice as common as yes/no questions in our test battery. Other common errors included polarity errors (e.g., *Can [for can't] she drink the the tea?*) and subject errors (e.g., *Can't you [for Can't she] drink the the tea?*). As a partial solution to this problem, statistical analyses were conducted on the *proportion* of each child's utterances

that constituted correctly formed questions (or auxiliary-placement errors etc. . . .) as a function of the total number of trials (out of two) in which the child produced some form of the target question (i.e., excluding those coded as *non-target* but not those coded as *unclassified*). Since two trials were conducted for each of the 12 question types, this means that a score of 0, 0.5 or 1 was recorded for each dependent measure, for each of the 12 question types, for each child.

## References

- Ambridge, Ben, Caroline F. Rowland, and Julian M. Pine  
2008 Is structure dependence an innate constraint? New experimental evidence from children's complex-question production. *Cognitive Science* 32(1), 222–255.
- Ambridge, Ben, Caroline F. Rowland, Anna L. Theakston, and Michael Tomasello  
2006 Comparing different accounts of non-inversion errors in children's non-subject wh-questions: 'What experimental data can tell us?' *Journal of Child Language* 33(3), 519–557.
- Ambridge, Ben and Julian M. Pine  
2006 Testing the Agreement/Tense Omission Model using an elicited imitation paradigm. *Journal of Child Language* 33(4), 879–898.
- Bannard, Colin and Danielle, E. Matthews  
2008 Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19(3), 241–248.
- Bowerman, Melissa  
1973 Structural relationships in children's utterances: Syntactic or semantic? In Moore, T. E. (ed.), *Cognitive Development and the Acquisition of Language*. New York: Academic Press, 197–213.
- Bowerman, Melissa  
1976 Semantic factors in the acquisition of rules for word use and sentence construction. In Morehead, D. and A. E. Morehead (eds.), *Normal and Deficient Child Language*. Baltimore, MD: University Park Press.
- Braine, Martin, D. S.  
1976 Children's first word combinations. *Monographs of the Society for Research in Child Development* 41(164), 1–104.
- Bybee, Joan  
1995 Regular morphology and the lexicon. *Language and Cognitive Processes* 10, 425–455.
- Bybee, Joan and J. Scheibman  
1999 The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37, 575–596.
- Dąbrowska, Ewa and E. Lieven  
2005 Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16(3), 437–474.
- DeVilliers, Jill  
1991 Why question? In Maxfield, T. L. and B. Plunkett. (eds.), *Papers in the Acquisition of wh: Proceedings of the UMASS Roundtable, May 1990*. Amherst MA: University of Massachusetts Occasional Papers.



- Erreich, Anne  
1984 Learning how to ask: Patterns of inversion in yes-no and wh-questions. *Journal of Child Language* 11, 579–592.
- Fletcher, Paul  
1985 *A Child's Learning of English*. Oxford: Blackwell.
- Guasti, Maria T., Rosalind Thornton, and Kenneth Wexler  
1995 Negation in Children's questions: The case of English. *Proceedings of the 19th Annual Boston University Conference on Language Development*, 228–239.
- Hattori, Ryoko  
2003 Why do children say did you went?: The role of do-support. In Brugos, Alejn, Linnea Micciulla, and Christine E. Smith (eds), *Proceedings of the 28th Boston University Conference on Language Development*. Boston, USA: Cascadilla Press.
- Hurford, James  
1975 A child and the English question formation rule. *Journal of Child Language* 2, 299–301.
- Kuczaj, Stanley. A. and Nancy Brannick  
1979 Children's use of the wh question modal auxiliary placement rule. *Journal of Experimental Child Psychology* 28, 43–67.
- Labov, William and Theresa Labov  
1978 Learning the syntax of questions. In Campbell, R. and P. Smith (eds.), *Recent Advances in the Psychology of Language*. New York: Plenum Press.
- Lieven, Elena V. M., Julian. M. Pine, and Gillian Baldwin  
1997 Lexically-based learning and early grammatical development. *Journal of Child Language* 24(1), 187–219.
- Lieven, Elena V. M, Heike Behrens, Jennifer Speares, and Michael Tomasello  
2003 Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30(2), 333–370.
- Maratsos, Michael and Stanley Kuczaj  
1977 Against the transformationalist account: A simpler analysis of auxiliary overmarking. *Journal of Child Language* 5, 337–345.
- Marchmann, V. A., B. Wulfeck, and S. E. Weismer  
1999 Morphological productivity in children with normal language and SLI: A study of the English Past Tense. *Journal of Speech, Language and Hearing Research* 42, 206–219.
- Rowland, Caroline F.  
2007 Explaining errors in children's questions. *Cognition* 104(1), 106–134
- Rowland, Caroline, F. and Julian, M. Pine  
2000 Subject-auxiliary auxiliary-placement errors and wh-question acquisition: 'What children do know?' *Journal of Child Language* 27(1), 157–181.
- Rowland, Caroline F. and Julian, M. Pine  
2003 The development of inversion in wh-questions: A reply to Van Valin. *Journal of Child Language* 30, 197–212.
- Rubino, R. and Julian, M. Pine  
1998 Subject-verb agreement in Brazilian Portuguese: What low error rates hide. *Journal of Child Language* 25, 35–60.

- Santelmann, Lynne, Stephanie Berk, Jennifer Austin, Shamitha Somashekar, and Barbara Lust  
2002 Continuity and development in the acquisition of inversion in yes/no questions: Dissociating movement and inflection. *Journal of Child Language* 29, 813–842.
- Savage, Ceri, Lieven, Elena, V. M., Anna L. Theakston, and Michael Tomasello  
2003 Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science* 6(5), 557–567.
- Stromswold, Karin  
1990 Learnability and the acquisition of auxiliaries. Unpublished Ph.D. dissertation, MIT.
- Theakston, Anna L., Elena V. M. Lieven, Julian M. Pine, and Caroline F. Rowland  
2000 The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language* 28(1), 127–152.
- Tomasello, Michael  
2003 *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tyack, David and Dorothy Ingram  
1977 Children's production and comprehension of questions. *Journal of Child Language* 4(2), 211–224.
- Valian, Virginia., and Lyman Casey  
2003 Young children's acquisition of wh-questions: The role of structured input. *Journal of Child Language* 30, 117–143.