J. Child Lang. **30** (2003), 609–635.
DOI: 10.1017/S0305000903005695 Printed in the United Kingdom

Determinants of acquisition order in wh-questions: re-evaluating the role of caregiver speech*

CAROLINE F. ROWLAND
University of Liverpool

JULIAN M. PINE
University of Nottingham

ELENA V. M. LIEVEN
Max Planck Institute for Evolutionary Anthropology, Leipzig

AND

ANNA L. THEAKSTON
University of Manchester

(*Received* 17 *April* 2002. *Revised* 18 *January* 2003)

ABSTRACT

Accounts that specify semantic and/or syntactic complexity as the primary determinant of the order in which children acquire particular words or grammatical constructions have been highly influential in the literature on question acquisition. One explanation of wh-question acquisition in particular suggests that the order in which English speaking children acquire wh-questions is determined by two interlocking linguistic factors; the syntactic function of the wh-word that heads the question and the semantic generality (or 'lightness') of the main verb (Bloom, Merkin & Wootten, 1982; Bloom, 1991). Another more recent view, however, is that acquisition is influenced by the relative frequency with which children hear particular wh-words and verbs in their input (e.g. Rowland & Pine, 2000). In the present study over 300 hours of naturalistic data from twelve two- to three-year-old children and their mothers were analysed in order to assess the relative contribution of

[*] We would like to thank all the families who took part in the research reported here. The collection and transcription of the data presented here was funded by the Economic and Social Research Council, Grant Number R000236393. Address for correspondence: Caroline Rowland, Department of Psychology, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool L69 7ZA, UK. tel: +44 (151) 794 1120; e-mail: crowland@liverpool.ac.uk

complexity and input frequency to wh-question acquisition. The analyses revealed, first, that the acquisition order of wh-questions could be predicted successfully from the frequency with which particular wh-words and verbs occurred in the children's input and, second, that syntactic and semantic complexity did not reliably predict acquisition once input frequency was taken into account. These results suggest that the relationship between acquisition and complexity may be a by-product of the high correlation between complexity and the frequency with which mothers use particular wh-words and verbs. We interpret the results in terms of a constructivist view of language acquisition.

INTRODUCTION

Many theories of language acquisition have argued for a role for semantic and/or syntactic complexity in children's acquisition of words and grammatical constructions. Some have claimed that children acquire semantically general verbs (verbs that encode very general meanings, e.g. *go*, *get*, *do*) more easily and thus more quickly than other more complex verbs (e.g. Clark, 1978; Pinker, 1989). Others (e.g. Blewitt, 1982) have argued that semantically simple size adjectives are acquired more easily than more complex adjectives. Similar propositions based around either semantic or syntactic complexity have been forwarded for morphology (e.g. Brown, 1973; Pinker, 1984), sentence types (e.g. Brown & Hanlon, 1970) and syntactic connectives (Bloom, Lahey, Hood, Lifter & Fiess, 1980).

Complexity's role in acquisition has been emphasized in the literature on wh-question development. The consensus seems to be that there is a relatively robust order of acquisition of wh-words in questions, in which the wh-words that encode syntactically simple relationships (e.g. *what* and *where*) are acquired before other wh-words that refer to more complex concepts (e.g. *why*, *how* and *when*). This robust sequence of acquisition in both production and comprehension has been reported for a variety of languages (for English see, for example, Tyack & Ingram, 1977; Bloom, Merkin & Wootten, 1982; also see Savić, 1975, for Serbo-Croatian; Clancy, 1989, for Korean; Forner, 1979, for German and Serbo-Croatian; Okubo, 1967, for Japanese; Wode, 1975, for German). Other studies have drawn upon the work on verb semantic generality and have identified an influence of the verb in wh-question acquisition. These studies have reported that early wh-questions tend to occur primarily with semantically general (or light) verbs and the copula, despite the fact that more complex verbs are produced at the same time in other structures (Johnson, 1981; Bloom *et al.*, 1982; Clancy, 1989).

One of the best-specified accounts of wh-question acquisition in terms of complexity is that proposed by Bloom and her associates (Bloom *et al.*, 1982; Bloom, 1991). The central proposition of this theory is that wh-question acquisition is determined by the syntactic and semantic complexity of the concepts encoded by the wh-words and the verbs to be acquired. It is suggested that the first wh-questions to emerge will be wh-identity questions; questions that ask for the identities of things or places. These are predicted to occur with what Bloom *et al.* termed the 'relatively simple' (Bloom *et al.*, 1982: 1086) wh-pronominals *what* and *where*, and should occur primarily with the copula. Later on, the wh-pronominals, which now also include *who*, are envisaged to start occurring with a greater variety of main verbs (e.g. *Where has he gone?*, *What are you doing?*). However, these verbs are expected to be restricted to what Bloom *et al.* termed pro-verbs such as *do* and *go* (also referred to as light verbs or semantically general verbs). These pro-verbs, or semantically general verbs, are said to be easier to acquire than other more descriptive verbs because they carry less information (e.g. *go* carries less information than *walk*), they involve fewer restrictions on the form of other sentence constituents (e.g. on the subject and object) and they are appropriate in a wider range of contexts (Bloom *et al.*, 1982).

Later still, the 'wh-sententials' (Bloom *et al.*, 1982: 1086) *when*, *how* and *why* are predicted to occur in the children's data, followed by the 'adjectival forms' (Bloom *et al.*, 1982: 1086) *which* and *whose*.[1] Bloom *et al.* consider the wh-sententials to be more complex than the wh-pronominals because 'the answers … specify a reason, a manner or a time that the entire event encoded in the sentence occurs' (Bloom *et al.*, 1982: 1086), a more complex operation than the simple referent identity function of the first wh-pronominal questions. Wh-adjectivals are last acquired because they are more complex still since they require the answer to 'specify something about an object constituent' (e.g. *Which ball?*, *Whose dinner?*; Bloom *et al.*, 1982: 1086). All these later acquired wh-words are also, it is argued, more likely to occur with descriptive verbs than the wh-pronominals; 'in sum, the children learned to ask wh-questions with descriptive verbs with those wh-question forms that were acquired late in the developmental sequence' (Bloom *et al.*, 1982: 1088).[2]

To summarize, according to this complexity account, the first wh-words – primarily *what* and *where* – should occur with the copula. Then these wh-words (together with *who*) should start to occur with semantically general

[1] These wh-forms are now termed wh-determiners. For consistency with Bloom *et al.*'s account we will continue to refer to these forms as wh-adjectivals. Thanks are due to Richard Ingham for pointing this out.

[2] There may be problems with Bloom *et al.*'s definitions of complexity. For example, *where* is considered a wh-pronominal but could easily be classified as a wh-sentential as, like *when*, *how* and *why*, it 'asks for information that pertains to the semantic relations among all the constituents in a sentence' (Bloom *et al.*, 1982: 1086). However, we have chosen not to discuss the issue of defining complexity in the present paper as it would involve far more discussion than the limits of the present paper allow. Other definitions of complexity will of course need to be tested in similar ways but those analyses are beyond the scope of the present paper.

verbs. Later still, wh-sententials will be acquired and, at about the same time, the children will start to use descriptive verbs, primarily with the wh-sententials. Finally, wh-adjectival forms will be acquired.[3]

On the face of it this account seems successfully to explain wh-question acquisition in English children. In particular, the data Bloom *et al.* present from seven children (aged 1;10 to 3;0) seem remarkably consistent with the acquisition order predicted from the interlocking influences of the syntactic complexity of the wh-word and the semantic generality of the verb. However, one problem is that the explanation does not consider the role of the speech that children hear; in particular, the frequency with which caregivers use particular wh-words and verbs. This is an important omission, as the complexity of a word and the frequency with which it is used are often strongly correlated. For example, the most frequently produced wh-words in German and Serbo-Croatian (Forner, 1979) and Korean (Clancy, 1989) are syntactically simple according to Bloom *et al.* (e.g. *what* and *where*) and the early learnt semantically general verbs in Ninio's (1996) data were also those that were the most frequent in both adult and child speech (Goldberg, 1999). What this means is that frequency, not complexity, may be the primary determinant of acquisition (Clancy, 1989). Thus, an alternative explanation of Bloom's data would state that children are producing wh-questions with particular wh-words and verbs that they have heard often in wh-questions in their caregiver's speech.

The role of the input in wh-question acquisition has not gone unchallenged. In particular, Savić (1975) has argued that the order in which wh-words start to appear in Serbo-Croatian speech to children, and the frequency with which they are used, does not correspond to the order and the frequency with which children use these questions. However, this conclusion seems to have been based on the fact that there is no very close match between child and adult data, rather than on statistical analysis. In fact, when we perform a correlation between the frequency of particular wh-word forms in the input and the order of acquisition of wh-words in children's speech based on the data Savić presented, the correlation is highly significant for both of the children studied ($r=0{\cdot}722$, $N=12$, $p=0{\cdot}008$ for Jasmina and $r=0{\cdot}763$, $N=12$, $p=0{\cdot}004$ for Danko).[4] Savić also reports a long period of

[3] Bloom *et al.* (1982) also consider the role of the linguistic contingency of the children's questions and the discourse adjustments that children make. They found little, or contradictory, effects of either discourse contingency or verb cohesion on acquisition for any questions but *why*. As a result, this part of the explanation will not be considered in the present paper.

[4] Forner (1979) performed correlations on Savic's data and reached a similar conclusion. However, her correlations yield slightly different results because she included all questions in her analysis, whereas in the present paper we have focused only on the wh-questions.

'incubation' in the first stages of question acquisition, so that a child's first use of a wh-word tends to appear months after it started to appear in the input. This is precisely what we would expect if cumulative input frequency was having an effect on order of acquisition.

In addition, more recent studies have demonstrated that acquisition is more heavily influenced by the frequency statistics of the speech that children hear than has previously been suggested. Computer models trained on input that captures the distributional characteristics of naturalistic language can 'learn' inflectional morphology and word class categories (see e.g. Finch, Chater & Redington, 1995; Cartwright & Brent, 1997). Similarly, some naturalistic and experimental studies have reported a role for the input in the acquisition of lexical categories and grammatical rules (see, for example, Goldberg, 1999). Even more significantly, some studies have found that the relative frequency of different constructions in the input correlates with the relative order of acquisition of those constructions in the child's speech (e.g. see Moerk, 1980, for morphology and the acquisition of specific prepositional phrases (though also see Pinker, 1981); Forner, 1979, for morphology; Naigles & Hoff-Ginsberg, 1998, for verbs; Hsieh, Leonard & Swanson, 1999, for the acquisition of plural noun and third singular verb inflections; Theakston, Lieven, Pine & Rowland, 2001, for verbs).

We now have two different explanations of wh-question acquisition. The first, proposed by Bloom *et al.*, suggests that the order in which wh-questions are acquired is determined by the syntactic and semantic complexity of the wh-words and verbs used in the question. The second, proposed here (and for wh-words only in Clancy, 1989), is that children acquire high frequency wh-words and verbs earlier than lower frequency lexemes. Complexity and input frequency are highly correlated themselves (e.g. Clancy, 1989), which means that both predictors will be similarly correlated with acquisition. In order to distinguish between them, we need to determine the relative contributions of each on acquisition.

The aim of the present paper was to establish the extent to which wh-complexity, verb semantic generality and input frequency predict the order of acquisition of wh-questions in children's early speech by investigating the data at the level of the individual wh-word + verb combination. At the level of the lexical item it may be possible to distinguish between the two highly correlated variables of complexity and frequency. According to the input explanation different forms of same verb (e.g. *is*, *are*) will be acquired at different times if their input frequencies differ. So if *what are* occurs less often in the input than *what is*, it will be acquired later. However, according to Bloom *et al.*'s account, there is no reason why different forms of the same verb, occurring with the same wh-word, should be acquired later than others – so *what are* should be acquired at roughly the same time as *what is*. To give another example, since two different forms of a semantically general verb (e.g. *go* and *going*) should be equally easy to acquire, there is no principled reason based on semantic generality why they should be acquired at different times.

The complexity account could, of course, incorporate a role for other influences such as tense and salience, which could affect the acquisition of two forms of the same verb such as *is* and *are*. In addition, there are other reasons why some verb forms may be more complex than others (e.g. why *are* may be more complex than *is*). However, these are, importantly, not predictions from Bloom's *et al.*'s linguistic complexity account. Although Bloom *et al.*'s account does not argue that the types of complexity discussed are the only predictor of acquisition, it does carry the implication that these types of complexity have the greatest effect on acquisition. It is this proposal, rather than the complexity account in its entirety, that the present analyses were designed to explain.

In order to achieve its aim, the present study had three objectives. First, two analyses established whether the order of acquisition of wh-words and verbs reported by Bloom *et al.* could be replicated on a new sample of early multi-word speech data. Second, the relative contributions of complexity and frequency to acquisition were investigated. This was achieved using a regression analysis that determined the power of the two predictors and also by investigating whether semantic generality and wh-complexity could explain order of acquisition effects that input frequency could not explain. Third, the paper aimed to discover whether any correlations that existed between input frequency and order of acquisition were due to a robust relationship between order of acquisition and frequency or were a by-product of correlations between mother and child frequency. Child frequency of use and order of acquisition are themselves highly correlated; thus correlations between order of acquisition and input frequency could simply reflect correlations between mother frequency and child frequency. The final analysis investigated this possibility.

METHOD

Participants

The participants were twelve children who took part in a longitudinal study of development. Six were from Nottingham, England and six from Manchester, England. Predominantly from middle-class backgrounds, the children were recruited through local nurseries, doctors' surgeries and newspaper advertisements on the basis that they were just beginning to produce multi-word speech (as measured by the MacArthur Communicative Development Inventory (Children)). All were monolingual English-speaking first born children whose mothers were the primary caregivers. Ages ranged from 1;8.22 to 2;0.25 at the start and 2;9.10 to 3;0.10 at the end of the study

https://doi.org/10.1017/S0305000903005695 Published online by Cambridge University Press

TABLE 1. *Age range, MLU range and total number of wh-questions asked by each of the 12 children*

Child	Age range	MLU range	Total no. wh-questions
Anne	1;10.7–2;9.10	1·61–3·46	593
Aran	1;11.12–2;10.28	1·41–3·84	364
Becky	2;0.7–2;11.15	1·46–3·24	1064
Carl	1;8.22–2;8.15	2·17–3·93	694
Dominic	1;10.24–2;10.16	1·20–2·85	153
Gail	1;11.27–2;11.12	1·76–3·42	499
Joel	1;11.1–2;10.11	1·33–3·32	379
John	1;11.15–2;10.24	2·22–2·93	182
Liz	1;11.9–2;10.18	1·35–4·12	412
Nicole	2;0.25–3;0.10	1·06–3·26	202
Ruth	1;11.15–2;11.21	1·41–3·35	81
Warren	1;10.06–2;9.20	2·01–4·12	287
Mean	—	1·58–3·49	409·17

(see Table 1). MLUs ranged from 1·06 to 2·22 at the beginning and 2·85 to 4·12 at the end of the study. The corpus is available on the CHILDES database (MacWhinney, 2000) and is referred to as the Manchester corpus (Theakston, Lieven, Pine & Rowland, 2000).

Procedure

The twelve children were audio-recorded in their homes for two separate hours in every three weeks for a year. The children engaged in everyday play activities with their mothers, half the time with their own toys and half the time with toys provided by the investigator. The data were orthographically transcribed using the CHILDES system and the age and MLU of the children at each tape calculated.

Speech corpora

Children. All spontaneous, complete, matrix wh-questions were extracted from the child's data. We excluded partially intelligible or incomplete utterances, utterances with parts marked as unclear, quoted utterances and routines (e.g. counting, nursery rhymes and songs). Full or partial repetitions or imitations of one of the previous five utterances were also excluded.

In order to replicate Bloom *et al.*'s main analyses, we extracted only those wh-questions that occurred with a main verb or a copula (e.g. *What are you doing?*, *Where've you gone?*, *Where's that?*), excluding those with an omitted main verb or copula. Other wh-question errors such as auxiliary omission errors, case errors and agreement errors were included.

The data were then divided into stages based on MLU according to Brown's (1973) criteria (see Table 2). At stage I, MLU ranged from 1·00 to

https://doi.org/10.1017/S0305000903005695 Published online by Cambridge University Press

TABLE 2. *Transcript numbers and MLU of first and last tapes for each child at each stage*

Child	Stage I (MLU 1–1·99)	Stage II (MLU 2–2·49)	Stage III (MLU 2·5–2·99)	Stage IV/V (MLU 3+)
Anne (tape nos.)	1–6	7–10	11–26	27–34
MLU range	1·61–1·92	2·27–2·21	2·62–2·97	3·09–3·46
Aran (tape nos.)	1–3	4–8	9–16	17–34
MLU range	1·41–1·83	2·22–2·27	2·57–2·97	3·08–3·84
Becky (tape nos.)	1–8	9–11	12–17	18–34
MLU range	1·46–1·97	2·06–2·41	2·50–2·90	3·26–3·24
Carl (tape nos.)		1–12	13–16	17–34
MLU range		2·17–2·49	2·70–2·75	3·07–3·93
Dominic (tape nos.)	1–10	11–21	22–34	
MLU range	1·20–1·78	2·12–2·48	2·87–2·85	
Gail (tape nos.)	1–3	4–8	9–24	25–34
MLU range	1·76–1·88	2·04–2·42	2·63–2·78	3·61–3·42
Joel (tape nos.)	1–8	9–16	17–34	
MLU range	1·33–1·87	2·00–2·48	2·56–3·32	
John (tape nos.)		1–24	25	26–34
MLU range		2·22–2·48	2·99	3·14–2·93
Liz (tape nos.)	1–5	6–12	13–18	19–34
MLU range	1·35–1·88	2·02–2·42	2·67–2·69	3·07–4·12
Nicole (tape nos.)	1–17	18–34		
MLU range	1·06–1·71	2·04–3·26		
Ruth (tape nos.)	1–12	13–25	26–30	31–34
MLU range	1·41–1·97	2·04–2·03	2·81–2·88	3·28–3·35
Warren (tape nos.)	1–2	3–6	7–11	12–34
MLU range	2·01–1·95	2·36–2·33	2·56–2·79	3·15–4·12

1·99; at stage II, MLU ranged from 2·00 to 2·49 and at stage III, MLU ranged from 2·50–2·99. Tapes for which the MLU was 3·00 or above were placed in stage IV/V. A child was regarded as moving to the next stage of development when three consecutive transcripts had MLUs over the MLU boundary.

Mother's data. All spontaneous, complete, matrix wh-questions were extracted from the mothers' data across all 34 transcripts. As with the children's speech, we included only wh-questions that occurred with verbs or the copula. The analysis was conducted on tokens as we wanted to extract a frequency count of the number of times a child heard a wh-word and verb together in a wh-question.

In order to ensure that the mothers' use of wh-words was independent of the effects of the children's use, pairwise correlations were calculated between the mothers for all the eight wh-words. All correlations were above 0·953 ($p<0{\cdot}001$), suggesting that the frequency of use was consistent across mothers and not influenced extensively by the individual children's use. In addition, we correlated the frequency of use of wh-words on the first and the last four tapes in the speech of individual mothers. This demonstrated that the relative frequency of wh-word use remained consistent over time despite the child's obvious improvement in language ability; all 12 correlations were above 0·92 ($p \leqslant 0{\cdot}002$).

Similar analyses have already been conducted on verb use in the Manchester corpus (see Theakston *et al.*, in submission), which demonstrated that the frequency with which mothers use verbs was similar across mothers (all correlations conducted were above 0·80, $p < 0{\cdot}01$ for the 10 children studied) and that the frequency of use of verbs did not change over time (correlations between verb use during the first and last four tapes were all above 0·90, $p < 0{\cdot}01$). Thus, we can be fairly confident that the mother's data was not dramatically affected by discourse requirements or the wh-word and verb use of individual children.

RESULTS

Order of acquisition of wh-words

The first analysis investigated whether the order of acquisition of wh-words in our corpus replicated that found by Bloom *et al*. Categorization of wh-words followed that suggested by Bloom *et al.*; *what*, *where* and *who* (the wh-pronominals) are the simplest, followed by *when*, *why* and *how* (the wh-sententials) and finally *which* and *whose* (the wh-adjectivals), which are considered the most complex. The acquisition order of wh-words was calculated for each of the twelve children. The 3rd use acquisition criteria adopted by Bloom *et al*. was used:[5] a wh-word was considered acquired when it had occurred in three different wh-questions (i.e. questions in which at least one element of the main clause differed e.g. *What is Jack doing?*, *What is he doing?*, *What is Jack eating?* or *Where is Jack going*, *Where has Jack gone*, *Where does he go?*).

Table 3 demonstrates the results. Although many of the wh-words were not acquired by all of the children, the order of acquisition was relatively consistent across the children and broadly mimicked the predicted order of wh-pronominals occurring before wh-sententials, which occurred before wh-adjectivals. In particular, *where* and *what* were the first acquired wh-words for all 12 children. In fact, wh-pronominals accounted for all the wh-questions produced by 11 of the children during stage 1 and 2 (Nicole also produced *why* in stage 2). There were some deviations from the predicted order of acquisition but few consistent differences across the 12 children. Four children acquired *how* or *why* before *who* and four acquired one of the wh-adjectivals (*which/whose*) before one or more of the wh-sententials (*why/how/when*). The complexity theory cannot explain these individual

[5] For one analysis (specified later) a first use acquisition criteria was used.

TABLE 3. *Order of acquisition of wh-words*

Child	Stage I	Stage II	Stage III	Stage IV & V	Not acquired
Anne	where (2) what (3)		who (12)	which (27)	**why/how/when/** whose
Aran		where (5)	what (12) who (13)	why (24) when (26) how (29) which (32)	whose
Becky	what (3) where (6)	who (11)	how (15)	why (18) **which** (32)	**when**/whose
Carl		what/where (1)			who/how/why/ which/whose/when
Dominic	what (4)	where (12)			who/how/why/ when/which/whose
Gail	what (1)	where (5)	**how** (16) **who** (20)		why/when/which/ whose
Joel		what (9) where (10)	**how** (17) **who** (18) **whose** (28)		**why/when**/which
John		what/where (2)			who/how/when/ why/which/whose
Liz	what/where (4)	who (9)		**which** (29)	**how/why/when**/ whose
Nicole	where (11)	what (18) who (23) why (31)			how/when/which/ whose
Ruth			what (26) **why** (28) where (30)	**who** (31)	how/which/whose/ when
Warren	where (1)			what (17) who (21)	which/why/how/ when/whose

Tape number at which verb was acquired is marked in parentheses. Wh-words that are not in the order predicted by Bloom *et al.* are in bold.

differences but, given the constraints of working with a sample of data, we would argue that the acquisition order fits quite closely onto that predicted.

Order of acquisition of verbs

The second analysis tested whether the Manchester corpus children acquired verbs in wh-questions in the predicted order (copula→semantically general→descriptive verbs). Unfortunately, categorizing verbs according to complexity was more difficult than categorizing wh-words. Although Bloom *et al.* discuss why pro-verbs may be relatively easy to acquire, they provide no definitive, objective criteria by which to define such verbs. The verbs *do*, *go* and *happen* are mentioned as pro-verbs but there is no indication of whether these are examples or should be treated as a definitive list. We have chosen to treat them as examples as we can see no reason to exclude other similar verbs, such as *come*, which also function as 'general, all-purpose' verbs (Bloom *et al.*, 1982: 1087).

As a result, in the present paper we decided to define semantically general verbs according to the criteria outlined in Theakston *et al.* (in submission), scheme 1, which states that a verb is considered semantically general if it has been classified as such by at least two of three influential studies concerned with semantic generality (Clark, 1978; Pinker, 1989; Ninio, 1999). In these papers, the researchers attempt to define semantically general verbs and suggest ways in which they may provide children with a route into syntax. Given that there are disputes as to how best to define verb semantic generality, we felt that a criterion based on a consensus among researchers rather than one scheme only would best satisfy the requirements of objectivity. The criterion is fairly restrictive (only *bring*, *come*, *do*, *get*, *give*, *go*, *make*, *put* and *take* are included) but more generous criteria tend to include all or most of the verbs produced by children in the early years, which reduces the predictive power of the theory (see Theakston *et al.*, in submission, for a discussion of this issue).[6]

The acquisition order of verbs in wh-questions was calculated from the data. A verb was said to be acquired when it had occurred in one or more of its inflectional forms in three different wh-questions (e.g. *Where are you going?*, *Where is he going?*, *Where has he gone?*→verb *to go* acquired). The results are presented in Table 4.

As predicted, the copula was acquired very early; it was the first acquired verb for nine of the 12 children and the second acquired verb for two of the other three children. Only Ruth acquired the copula late. Other early-acquired verbs (acquired in stages I and II) were predominantly semantically general verbs (except *fit*, acquired in stage 1 by Anne, *want*, acquired by Becky in stage II, *happen*, acquired in stage 2 by Carl and John, and *bought*, acquired in Stage II by Nicole). At stages III and IV/V many more descriptive verbs were acquired, as predicted by the complexity account.

The relative contributions of semantic generality, wh-complexity and input frequency to acquisition

The third analysis investigated whether wh-complexity and verb semantic generality have an effect once we take input frequency into account. For the

[6] The verb *happen* is specified by Bloom *et al.* as a pro-verb but does not meet the criteria for the present classification scheme. In order to maintain the objectivity of the scheme, *happen* is not categorized as a semantically general verb. However, alternative analyses have been conducted and including *happen* as a semantically general verb does not affect the pattern of results reported here.

TABLE 4. *Verbs acquired at each stage*

	Stage			
Child	I	II	III	IV/V
Anne	Copula Be (1) Go (3) **Fit (4)** Do (6)		**Like (14)** **Sit (17)**	Get (29) **Want (31)** Put (32) **Say (32)** **Have (34)**
Aran		Go (5)	Copula Be (9) **Eat (15)**	Do (18) **Buy (18)** Get (19) **Call (21)** Make (26) **Drink (26)** Come (26) Take (27) Put (29) **Say (29)** **Happen (32)**
Becky	Copula Be (2) Do (7) Go (7)	**Want (11)**	Get (13) **Call (13)** Make (14) **Like (17)**	**Bring (18)** **Find (18)** **Feel (19)** **Live (22)** **Play (24)** Put (25) **Sit (26)** **Cut (26)** **Buy (28)** **Say (28)** **Happen (29)** Have (29) **Hide (29)** **Eat (30)** **Break (30)** Come (32) **Tie (32)** **Build (34)**
Carl		Copula Be (1) Do (1) Go (1) **Happen (1)**		Come (20) **Want (34)** **Find (34)**
Dominic	Copula Be (4)	Go (12)		
Gail	Copula Be (1)	Go (5)	Do (13) **Call (14)** Get (14) Come (20)	Put (28) **Play (30)** **Like (31)** **Say (32)** **Have (33)**
Joel	Copula Be (7)	Go (12) Get (16)	Do (17) Come (25) Put (29)	

https://doi.org/10.1017/S0305000903005695 Published online by Cambridge University Press

TABLE 4. (*Cont.*)

Child	Stage I	II	III	IV/V
			Want (30) **Call (34)**	
John		Copula Be (2) Go (4) Do (9) **Happen (24)**		
Liz	Copula Be (3) Go (4)	Do (10)		Get (22) **Like (29)** **Want (34)**
Nicole	Copula Be (14) Go (15)	Get (25) Do (25) **Bought (26)** Put (33)		
Ruth			Do (26) Go (28) **Want (29)**	**Look (31)** Copula Be (31) Get (33)
Warren	Go (1)	Copula Be (4)		Do (19) **Happen (20)** Get (25) Make (28)

Tape number at which verb was acquired is marked in parentheses. Verbs marked in bold are descriptive verbs.

children's data we calculated the order of acquisition of all the individual wh-word + main verb (wh + verb) combinations used by the 12 children (e.g. *where* + *going*, *what* + *do*) and recorded the number of the transcript at which they were acquired.[7] Wh + verb combinations were considered acquired at 1st use rather than at 3rd use because of a lack of rich data for many of the children; for nine of the children only wh-pronominals were considered acquired by the 3rd use criterion. Third use criterion results are provided in footnotes for the three children for whom these were available.

To confirm the reliability of the 1st use acquisition criterion, we correlated the order of acquisition of wh + verb combinations according to the two criteria. This was only possible for nine of the children (since Dominic only acquired 3, John only acquired 7 and Ruth only acquired 4 wh + verb combinations according to the 3rd use criterion). Despite a paucity of data,

[7] Since the present analyses do not rely on statistical comparisons between children, it was not necessary to use MLU to ensure uniformity of developmental level across children. In fact, MLU would have been an unreliable measure of order of acquisition as the children's MLUs did not uniformly increase from one transcript to the next (e.g. John's MLU was 3·24 at transcript 27 but 3·04 at transcript 28).

https://doi.org/10.1017/S0305000903005695 Published online by Cambridge University Press

TABLE 5. *Intercorrelations between input frequency, semantic generality and wh-complexity with order of acquisition*

	N	Order × frequency (*r*)	Order × semantic-generality (*r*)	Order × wh-complexity (*r*)	Order × wh-complexity + semantic generality combined (*r*)
Anne	66	0·564**	0·124	0·442**	0·385**
Aran	97	0·493**	0·120	0·326**	0·284**
Becky	144	0·419**	0·247**	0·263**	0·370**
Carl	35	0·369*	0·075	0·321	0·248
Dominic	13	0·604*	0·448	0·328	0·615*
Gail	44	0·513**	0·015	0·190	0·110
Joel	64	0·470**	0·262*	0·124	0·344**
John	17	0·421	0·395	0·107	0·446
Liz	32	0·468**	0·288	0·413*	0·419*
Nicole	40	0·532**	0·481**	0·363*	0·575**
Ruth	28	0·058	0·205	0·189	0·066
Warren	30	0·474**	0·436*	0·138	0·397*

* = sig. at 0·05; ** = sig. at 0·01.

all correlations were above 0·62 (range $r = 0{\cdot}62–0{\cdot}93$, range $n = 10–90$, range $p < 0{\cdot}001–0{\cdot}014$), and all were significant at 0·05 (eight were significant at 0·01).

Each wh + verb combination produced by each child was then categorized for whether it occurred with a wh-pronominal (categorized as 1), a wh-sentential (2) or a wh-adjectival (3) form, and then was categorized again, separately, for whether it occurred with a copula form (1), a semantically general verb (2) or a descriptive verb (3). There were, thus, two scores for each wh + verb combination produced by each child; one for wh-complexity and one for semantic generality. For each wh + verb combination produced by a child, the frequency of these combinations in that child's input data was also calculated. Order of acquisition was then correlated with semantic generality, wh-complexity and input frequency (see Table 5).

Semantic generality correlated significantly with order of acquisition for only four of the children (i.e. the semantic generality of the verb in each wh + verb combination did not seem to account for the acquisition order of the wh + verb combinations for eight of the children). Wh-complexity correlated significantly with order of acquisition for only five of the 12 children. However, there were significant correlations between order of acquisition and input frequency for 10 of the 12 children.

A regression analysis was then conducted to establish the relative contributions of semantic generality, wh-complexity and input frequency independently of the effects of the other two predictors.[8] Again, the results

[8] Traditionally, regression analyses are only performed when there are a certain number of observations per predictor but the number of observations required is in dispute. Some

TABLE 6. *Simultaneous regression predicting order of acquisition with frequency of wh+verb in mother speech, verb semantic generality and wh-word syntactic complexity as predictors*

	R^2	Adjusted R^2	Frequency (beta)	Semantic generality (beta)	Wh-complexity (beta)
Anne	0·424**	0·396	0·488**	0·008	0·333**
Aran	0·292**	0·269	0·455**	0·038	0·222*
Becky	0·238**	0·222	0·333**	0·162*	0·221**
Carl	0·216	0·140	0·372 ($p=0·052$)	0·083	0·264
Dominic	0·491	0·321	0·369	0·353	0·334
Gail	0·275**	0·220	0·503**	0·045	0·090
Joel	0·249**	0·212	0·393**	0·185	0·129
John	0·225	0·046	0·179	0·331	0·176
Liz	0·343**	0·273	0·391*	0·108	0·313
Nicole	0·428**	0·380	0·347*	0·313*	0·256
Ruth	0·117	0·006	0·141	0·322	0·225
Warren	0·271*	0·187	0·355	0·189	0·148

* = sig. at 0·05; ** = sig. at 0·01.

appear to favour the input frequency predictor over the complexity predictors (see Table 6).

All together, the variables under consideration had a significant effect on acquisition for eight of the 12 children. However, semantic generality only contributed to order of acquisition for two of the 12 children and wh-complexity contributed to acquisition for only three of the children. Input frequency had a significant effect on order of acquisition at a significance value of 0·05 for seven of the children (for one additional child, Carl, the significant value was at 0·052 (Carl frequency beta = 0·04)).[9]

statisticians have argued for a more stringent rule (e.g. at least 50 observations for every 1 predictor) and others for more lenient measures (e.g. 5 or 10 observations for every 1 predictor). In the present study, even according to the relatively lenient 10:1 rule, there should be at least 30 observations for each child, which would mean we would exclude John, Ruth and Gail's data. According to the 50:1 rule we should have 150 observations per analysis, which would leave no data available for analysis. After consultation with a statistician, we decided that incorporating a stringent criterion into the present study would simply reduce the number of children included while having very little effect on the patterning of results and the conclusions drawn. As a result, the analyses will be performed on all 12 children's data. It must be noted though that strictly speaking (even according to the lenient 5:1 rule) Dominic's data should be excluded from Table 6 (though not from Tables 7 and 10 which investigate only 2 predictors and thus require only 10 observations per analysis).

[9] For the three children for whom third use criterion results were available, two had significant correlations between frequency and acquisition (Becky $r=0·565$, $N=40$, $p<0·01$; Liz $r=0·674$, $N=13$, $p<0·05$) and two had significant correlations between semantic generality and acquisition (Gail $r=0·707$, $N=16$, $p<0·01$; Liz $r=0·703$, $N=13$, $p<0·01$). Wh-complexity was never significantly correlated with acquisition. The regression results revealed that only frequency was a significant predictor of

TABLE 7. *Simultaneous regression predicting order of acquisition with frequency of wh + verb in mother speech and verb semantic generality + wh-word syntactic complexity combined as predictors*

	R^2	Assisted R^2	Frequency (beta)	Wh-complexity + semantic generality combined (beta)
Anne	0·369**	0·349	0·493**	0·236*
Aran	0·252**	0·236	0·452**	0·101
Becky	0·231**	0·220	0·328**	0·252**
Carl	0·142	0·089	0·326 (sig. at 0·1)	0·089
Dominic	0·488*	0·385	0·388	0·412
Gail	0·263**	0·227	0·516**	0·014
Joel	0·249**	0·224	0·396**	0·181
John	0·219	0·108	0·205	0·298
Liz	0·318**	0·271	0·389*	0·324*
Nicole	0·426**	0·395	0·344*	0·422**
Ruth	0·015	0·065	0·117	0·123
Warren	0·235*	0·179	0·380	0·129

* = sig. at 0·05; ** = sig. at 0·01.

Wh-complexity and verb semantic generality are not significant predictors when considered separately, but it could be that their predictive power increases when they are combined. To establish if this was the case, the analyses were repeated with the wh-complexity and verb semantic generality variables combined (e.g. for *what is*: wh-complexity rating = 1, verb semantic generality rating = 1, therefore combined rating = 2). In these analyses too, input frequency was the more powerful predictor. The complexity variable only correlated with order of acquisition for eight of the 12 children (see Table 5) and for the nine children for whom the regression analysis resulted in significant results, complexity was a significant predictor for only four (see Table 7). However, there were significant correlations between order of acquisition and input frequency for 10 of the 12 children and, in the regression analysis, input frequency significantly predicted order of acquisition for seven of the nine children for whom the regression analysis produced significant results.[10]

acquisition for Becky (frequency beta = 0·512, $p < 0·05$) and only semantic generality was a significant predictor for Gail (generality beta = 0·605, $p < 0·05$). None of the factors contributed to acquisition for Liz. This mixed pattern of data can neither confirm not disconfirm Bloom *et al.*'s analysis. However, the fact that the samples for Gail and Liz were small ($N = 16$ for Gail and $N = 13$ for Liz) suggest that perhaps Becky's data is the more reliable. Further work on richer data is necessary.

[10] For the 3rd use criterion analysis, frequency was the only significant predictor of acquisition for two of the children (Becky frequency beta = 0·52, $p < 0·01$ and Liz frequency beta = 0·52, $p < 0·05$). Complexity was the only significant predictor for Gail (complexity beta = 0·56, $p < 0·05$).

https://doi.org/10.1017/S0305000903005695 Published online by Cambridge University Press

TABLE 8. *The mean frequency in the input of each of the combinations acquired by the children early on (MLU 1–2·49) and later on (MLU 2·50+)*

	Mean input frequency of combinations acquired at Stage I (MLU = 1·00–2·49)	Mean input frequency of combinations acquired at Stage II (MLU = 2·50+)
Anne	103·57	13·67
Aran	73·4	21·37
Becky	61·22	4·86
Carl	66·44	3·5
Dominic	220·17	51·71
Gail	116·9	16·24
Joel	86·79	18·16
John	61·5	21·6
Liz	40·5	8·8
Ruth	104·67	54·04
Warren	196·33	35·04
Mean	102·86	22·64

To summarize, when we focus on distinguishing between the predictions of the complexity explanation and the input explanation, we find that, although none of the variables were significant predictors for all 12 children, the frequency of particular wh+verb combinations in the input is a better predictor of the order of acquisition of wh-questions than either the wh-complexity of the wh-word or the semantic generality of the verb.

In fact, the difference in input frequency between early and late acquired combinations was substantial. In order to illustrate this, the children's data were divided into two stages based on Brown's (1973) stages. Stages I and II were combined and stages III and IV/V were combined as some children did not acquire any combinations at one particular stage. Then the mean input frequency of the combinations acquired at the early (MLU 1–2·49) and later stages (MLU 2·50+) was calculated. One child acquired no combinations at the later stages (Nicole) so is not included (see Table 8).

Over the eleven children, the mean input frequency of the early-acquired combinations was 102·86 (S.D. = 57·00), compared to a mean frequency of 22·63 (S.D. = 17·32) for the late acquired combinations. Thus, as we would expect from the correlations above, the early-acquired combinations were of much greater frequency in the input than those that were late acquired.

Do wh-complexity and semantic generality explain other acquisition effects?

The results so far suggest that some of the effects attributed to complexity in acquisition may be by-products of a high correlation between input frequency and complexity. However, input frequency does not account for

https://doi.org/10.1017/S0305000903005695 Published online by Cambridge University Press

all the order of acquisition effects (it accounts for only 13 to 36% of the variance across the 10 children for whom input frequency and order of acquisition were significantly correlated). In particular, there are some combinations that are early acquired by the children that are low frequency in our sample and there are some late-acquired combinations that are of high input frequency. It could be that frequency and complexity interact: some low frequency combinations are learnt early because they are linguistically simple and some high frequency combinations are learnt late because they are more complex.

We identified the wh + verb combinations that were early acquired but were of low frequency in the input (occurred fewer than 20 times) and those that were late acquired but high frequency (occurred more than 20 times). Then we calculated the mean semantic generality rating (where 1 = copula, 2 = semantically general and 3 = descriptive verb) and mean wh-complexity rating (where 1 = wh-pronominal, 2 = wh-sentential and 3 = wh-adjectival) of these combinations for each child. One child produced no Stage III and IV/V data so is not included in the analysis. The results can be seen in Table 9.

Three of the children produced no early-acquired but low frequency combinations and one of the children produced no late-acquired high frequency combinations. For the remaining seven children there was no evidence that low frequency combinations were acquired early because they were linguistically simple or that high frequency combinations were acquired late because they were complex. In fact, for six of the seven children, the mean complexity of the early-acquired wh-words was greater than that of the late acquired wh-words. Similarly, for five of the seven children, the verbs produced earlier on were less likely to be semantically general than the late-acquired verbs. It does not seem to be the case that complexity explains why some high frequency wh-words and verbs are learnt late and some low frequency items are learnt early.

Child frequency of use

The results suggest so far that input frequency is having some impact on order of acquisition. However, we need to be able to dismiss the possibility that we are not capturing order of acquisition at all but simply frequency of use. Since frequently used terms are more likely to occur in a sample than more rarely used terms, highly frequent items are more likely to occur earlier in a longitudinal sample of data than low frequency terms, even if both are acquired at the same time. This means that our order of acquisition measure could in fact simply be measuring child frequency of use and our correlation merely indicating that mothers and children tend to use the same wh-words and verbs. In order to control for this possibility, we counted the number of times that each child produced each wh-word + verb combination (in tokens)

TABLE 9. *Wh-complexity and semantic generality ratings of wh+verb combinations that were acquired early but of low input frequency and of wh+verb combinations acquired late but of high input frequency*

	Number of combinations		Early-acquired but low frequency wh+verb		Late-acquired but high frequency wh+verb	
Child	No. early-acquired low input frequency combinations/total no. early-acquired combinations	No. late-acquired high input frequency combinations/total no. late-acquired combinations	Mean wh-complexity rating	Mean semantic generality rating	Mean wh-complexity rating	Mean semantic generality rating
Anne	4/14	12/52	1·25	2·5	1	2·3
Aran	2/5	18/92	1·5	2·5	1	2
Becky	7/18	7/126	1·38	2·13	1	2·14
Carl	13/27	0/8	1·15	2·59	No high frequency wh+verb	
Dominic	0/6	3/7	No low frequency wh+verb		1·67	1·33
Gail	0/10	6/34	No low frequency wh+verb		1	2·33
Joel	7/14	10/50	1·29	2·14	1	2
John	4/12	2/5	1·25	2	1	1·5
Liz	5/12	2/20	1·2	2	1	2·5
Ruth	0/3	7/25	No low frequency wh+verb		1	2
Warren	1/3	14/27	1	2	1·07	1·79

Key

Wh-complexity: 1 = wh-pronominal
2 = wh-sentential
3 = wh-adjectival

Semantic generality: 1 = copula
2 = semantically general
3 = descriptive verb

Therefore, the higher the mean, the more complex the wh-word or verb.

TABLE 10. *Simultaneous regression predicting order of acquisition with frequency of wh+verb in child and mother speech as predictors*

	R^2	Adjusted R^2	Frequency in input (beta)	Frequency in child (beta)
Anne	0·328**	0·306	0·470**	0·134
Aran	0·264**	0·249	0·337*	0·214
Becky	0·217**	0·206	0·109	0·371**
Gail	0·291**	0·257	0·388*	0·210
Joel	0·267**	0·243	0·278 (p=0·064)	0·288 (p=0·055)
Liz	0·318**	0·271	0·385*	0·326*
Nicole	0·482**	0·454	0·057	0·652**

* = sig. at 0·05; ** = sig. at 0·01.

to give us a count of the frequency with which the children produced each combination. We then performed a regression, with frequency in the input and in the child's speech as predictors of order of acquisition, on the data for the seven children for whom input frequency was a significant predictor of order of acquisition (see Table 10).

The results are not clear-cut but do suggest that our correlations between order of acquisition and input frequency cannot be attributed solely to correlations between child and mother frequency of use. For four of the seven children, input frequency was still a significant predictor of order of acquisition when child frequency of use was removed, and for all four of these, input frequency was a more powerful predictor than frequency of use in the child. Frequency of use was correlated with order of acquisition for three of the children but was a more powerful predictor of order of acquisition than input frequency for only two of the children. Thus, although the results are mixed, they suggest that our regression analyses are not grossly confounded by child frequency of use.

DISCUSSION

The present study investigated the relative predictive power of two explanations of the order of acquisition of wh-questions; one based on linguistic complexity and one based on input frequency statistics. The study first verified that the order of acquisition of wh-words and verbs in a new corpus (the Manchester corpus) was consistent with that identified by Bloom *et al.* (1982) and then assessed the relative predictive power of complexity and input frequency on the order of acquisition of particular lexical wh-word and verb combinations. The results showed that of the two, input frequency was almost always the more powerful predictor of order of acquisition, even when complexity was taken into account. When the effect of input frequency was removed, semantic generality and wh-complexity were significant predictors for very few of the children. In fact, the differences in input frequency between the late acquired and early acquired wh + verb combinations were striking; the early acquired wh + verb combinations (acquired during stages I and II) occurred on average 102·86 times in the input, compared to the late acquired combinations (acquired during stages III and IV/V) that occurred on average only 22·63 times.

The analyses also showed that wh-complexity and verb semantic generality were not significant determinants of the acquisition effects that could not be attributed to input. For most children, early-acquired, low frequency wh-words were actually more complex than those that were late-acquired but high frequency. Similar effects for verbs were found. There was little significant evidence that semantic generality and wh-complexity had a strong and consistent effect on acquisition over and above the influence of input frequency.

The final analysis tested whether the order of acquisition data was confounded by the frequency with which children use particular wh-words and verbs. Although the results were not clear-cut, they did suggest a role for input frequency in order of acquisition for the majority of the children.

In summary, the apparent effects of wh-word syntactic and verb semantic complexity on the order in which children acquire wh-questions are, in the main, more clearly explained in terms of the frequency with which children hear particular wh-words and verbs in wh-questions than by complexity. It may be that children initially learn small scope semi-formulaic schemas based around the combinations of lexical items that have occurred relatively frequently in the input – input frequency focuses children's attention on particular high frequency wh-word + verb, wh-word + auxiliary + verb and/or wh-word + auxiliary combinations. Generalization will later occur as children start to learn more forms (including low frequency forms) and develop links between the words that share distributional (and perhaps semantic and phonological) properties. Incidentally, this account could also explain why children seem to be slower to learn verbs in wh-questions than in other contexts (Bloom *et al.*, 1982; Kuczaj, 1986). Children who learn a verb in a particular context (e.g. declaratives) will only be able to generalize that verb to other contexts (e.g. wh-questions) later in development, once they have built up patterns of association between verb slots in declaratives, yes–no and wh-questions. The development of an adult-like grammar could then involve a gradual broadening of scope as the children build up knowledge of the number of and the overlap between the small-scale schemas in which particular items can occur. Input frequency plays a part in constraining not only order of acquisition but also the development of broad syntactic categories and relations from lexically specific formulae (see, for example, Pine, Lieven & Rowland, 1998, for work on naturalistic data or see Redington & Chater, 1997; Seidenberg & MacDonald, 1999, for summaries of computer simulations using distributional analysis). The findings of the study are, thus, consistent with a constructivist account of language learning based on the idea that children 'must construct their grammatical categories on the basis of gradual learning of phonological, distributional and functional information embedded in the input' (Peters, 2001 : 236).

The account as presented here is over-simplified and focuses only on the distributional analysis part of a constructivist model (e.g. Goldberg, 1999). However, it may be the starting point around which a more complex theory of acquisition can be built. For example, apart from gross input frequency statistics, we could investigate other effects of the distribution, such as the syntactic diversity of the contexts in which a particular lexical item occurs or the combined frequencies of related verbs or wh-words. One particular modification that may be important is to take account of the difference between the type and token frequency of wh-questions. Modelling a wh-word repeatedly in the same wh-question may have different implications for learning than the modelling of a wh-word in a variety of different questions (see Bybee, 1995, for a consideration of the importance of type frequency and Naigles & Hoff-Ginsberg, 1998, for a similar explanation of verb learning).

In fact, it may be that such a distinction can explain some of the effects in our data that were not explained by the grosser frequency statistics of the input. For example, of the 43 early-acquired but low frequency wh+verb combinations identified in Table 9 above, 31 were produced in only one or two different question types. Five of these were errors of commission that would not have occurred in the input (*Where goes*, *What you wants*, *Where's choo+choo trains pushing*, *where's it my put it*, *Why it's all gone?*) and suggest some productivity. However, the remaining 26 wh+verb combinations could have been rote-learned questions or question fragments (see Table 11). Thus, although further work needs to be done on such forms, it may be that a more sophisticated version of the frequency account can explain more of the acquisition data.

A constructivist account of language acquisition also needs to assess the effect of non-distributional influences and how they interact with distributional information. Other effects will include, for example, phonological salience (Moerk, 1980), the child's communicative interests (Clancy, 1989) and cognitive complexity (e.g. Brown & Hanlon, 1979). It is also important to evaluate wh-question use in different contexts as the distributional characteristics of verbs and wh-words may differ in other settings (e.g. Hoff-Ginsberg, 1991). In fact, we may find some of these factors explaining why some high frequency wh+verb combinations are learnt late. For example, six of the children in the present study learnt copula *are* quite late on, despite the fact that it occurred with relatively high frequency in their input. It may be that copula *are* is of low phonological salience in English speech and is thus picked up late (see Moerk, 1980). Similarly, past tense

TABLE 11. *Early-acquired but low frequency combinations identified in Table 8 that occurred only once or twice in the children's data*

Child	Wh + verb	Wh-question
Anne	what + fit	what this fit
	how + doing	how do-ing
	where + goes	where go-es (error: tense marked on main verb)
Aran	why + crying	why boy cry-ing
Becky	what + happened	what happened (produced twice)
	what + wants	what you want-es (error: tense marked on main verb)
	how + do	how do you do (produced 3 times)
	which + want	which pattern do you want to have?
		which piece want?
Carl	what + bump	what bump my head?
		what bump head?
	what + happening	what-'is happen-ing?
	what + colouring	what colour-ing in?
	where + coming	where-'is daddy come-ing?
		where-'is gordon come-ing?
	where + hide	where hide?
	whose + are	whose clothes are these?
	where + pushing	where-'is choo + choo train-s push-ing? (error: from context, target is *what-'are choo + choo train-s pushing?*)
	where + find	where find it?
		where carl find?
	where + stand	where stand?
	what + going	what-'is go-ing on?
		what-'is go-ing on next?
Joel	how + got	how many got?
	what + going	what-'is go-ing on?
		what-'re we go-ing to do?
	what + need	what you need?
	what + found	what i found (produced twice)
	what + ringing	what-'is that ring-ing
	how + was	how old was Rachel
John	what + go	what-'s nut go? (from context, target seems to be *where's nut go*)
	how + are	how are you?
	what + happen	what happen the chip
Liz	where + put	where you put it?
		where-'is it my put it (error: unclear error form)
	what + happened	what happened
	what + are	what are those (produced 3 times)
	why + gone	why it-'has all gone? (error: auxiliary occurs after instead of before subject)

Only errors of commission are identified as errors. Omission errors (e.g. *what I found*) are not marked. Questions are coded morphologically according to CHAT convention.

https://doi.org/10.1017/S0305000903005695 Published online by Cambridge University Press

forms and the wh-word *when* may be delayed in acquisition until children master the cognitive notion of temporal relations (Brown & Hanlon, 1979); in our data, three children acquired *was* late despite its high frequency in their input.

Most importantly, the effect of linguistic complexity must be further assessed. Although the present study has concluded that input frequency is a more important predictor than complexity, we only considered complexity as defined by Bloom *et al.* (1982). There are other types of complexity to be assessed. In fact, at some minimal level complexity is bound to have an effect – a child's level of conceptual understanding must influence her understanding of the concepts expressed in the input and her ability to acquire the words to express these concepts (Slobin, 1973; Maratsos, 1979).

There may also be a more substantial role for complexity in other senses. For example, the unmarked form of a category (e.g. present, 1st person in English) may be less complex conceptually than the marked form (e.g. past tense) and may be more easily acquired. So, for example, the form *ran* would be expected to be acquired after the form *run*. Another possibility is that complexity may play a role in the sense that length of utterance can affect acquisition; wh-identity questions may be acquired first because they occur with the copula and are thus shorter and simpler than questions that require an auxiliary and a main verb.[11] These effects could interact with frequency; for example, once a child is ready to acquire the terms that express a concept (e.g. temporal relations) the order in which s/he acquires them could be influenced by their relative frequency. It may be that these types of interactions will allow us to account for significantly more of the variance than is explained by input frequency alone.

To conclude, the present study has demonstrated that although the relative input frequency of particular wh-word + verb combinations cannot be said to be the sole or even primary predictor of wh-question acquisition, it does seem to be is a more powerful predictor than linguistic complexity as defined by Bloom *et al.* (1982). We have suggested that such findings are compatible with some of the constructivist ideas that discuss a role for distributional analysis in language acquisition.

The role of the input has often been dismissed (e.g. Valian, 1999) because much of the research on the relationship between acquisition and the input has generally failed to find effects. However, many of the input analyses on which this conclusion is based are global analyses. They measure, for example, the length of parental utterances, how frequently parents produce declaratives and interrogatives, verbs, nouns and pronouns, or how frequently parents expand their children's utterances (see Valian, 1999, for

[11] Thanks are due to Edith Bavin and two anonymous reviewers for pointing out some of the ways in which complexity could affect acquisition.

a review of such studies). If, as we claim, children are analysing input on the level of the lexical item (e.g. building up knowledge of how specific words behave in low-scope lexical formulae), conducting analyses on these more abstract levels may obscure correlations between input frequency and order of acquisition. In addition, some of these analyses rely on finding individual differences between mothers, and searching for corresponding differences between children. This means that a lack of variation in both maternal and child data might conceal significant input–acquisition relationships. As we have attempted to show here, analyses performed at the level of the lexical item that do not rely on finding differences between mothers (e.g. that measure the relative frequency of occurrence of particular lexemes within, rather than across, mothers) may reveal interesting correlations between children's acquisition and the input.

At the very least, the effect of input frequency needs to be taken into account when we are building a model of a language acquisition mechanism, if only to distinguish between rote-learned forms, lexically specific schema and more productive utterances. What may be interesting is the effects input cannot explain but, in order to isolate these data, it is important initially to discover what can be attributed to the relative frequency of language structures in the input. There will be a role for linguistic complexity but its effect is different and perhaps more complex than has often previously been assumed.

REFERENCES

Blewitt, P. (1982). Word meaning acquisition in young children: a review of theory and research. In H. W. Reese (ed.), *Advances in child development and behavior* **17**. New York: Academic Press.

Bloom, L. (1991). *Language development from two to three*. Cambridge: CUP.

Bloom, L., Lahey, M., Hood, L., Lifter, K. & Fiess, K. (1980). Complex sentences: acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language* **7**, 235–61.

Bloom, L., Merkin, S. & Wootten, J. (1982). Wh-questions: linguistic factors that contribute to the sequence of acquisition. *Child Development* **53**, 1084–92.

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

Brown, R. & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (ed.), *Cognition and the development of language*. New York: Wiley.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* **10**(5), 425–55.

Cartwright, T. A. & Brent, M. R. (1997). Syntactic categorisation in early language acquisition: formalizing the role of distributional analysis. *Cognition* **63**, 121–70.

Clancy, P. (1989). Form and function in the acquisition of Korean wh-questions. *Journal of Child Language* **16**, 323–47.

Clark, E. V. (1978). Discovering what words can do. In D. Farkas, W. M. Jacobsen & K. W. Todrys (eds), *Papers from the paracession on the lexicon, Chicago Linguistics Society, April 14–15 1978*. Chicago: Chicago Linguistic Society, University of Chicago.

Finch, S. P., Chater, N. & Redington, M. (1995). Acquiring syntactic information from distributional statistics. In J. Levy, D. Bairaktaris, J. Bullinaria & P. Cairns (eds), *Connectionist models of memory and language*. London: UCL Press.

Forner, M. (1979). The mother as LAD: interaction between order and frequency of parental input and child production. In F. R. Eckman & A. J. Hastings (eds), *Studies in first and second language acquisition*. Rowley, MA: Newbury.

Goldberg, A. E. (1999). The emergence of the semantics of argument structure constructions. In B. MacWhinney (ed.), *The emergence of language*. Mahwah, NJ: Erlbaum.

Hoff-Ginsberg, E. (1991). Mother–child conversation in different social classes and communicative settings. *Child Development* **62**, 782–96.

Hsieh, L., Leonard, L. & Swanson, L. (1999). Some differences between English plural noun inflections and third singular verb inflections in the input: the contributions of frequency, sentence position and duration. *Journal of Child Language* **26**, 531–43.

Johnson, C. E. (1981). Children's questions and the discovery of interrogative syntax. Unpublished doctoral dissertation, Stanford University.

Kuczaj, S. (1986). General developmental patterns and individual differences in the acquisition of copula and auxiliary be forms. *First Language*, **6**(2), 111–17.

MacWhinney, B. (2000). *The CHILDES project: tools for analysing talk*. Mahwah, NJ: Erlbaum.

Maratsos, M. (1979). How to get from words to sentences. In D. Aaronson & R. W. Rieber (eds), *Psycholinguistic research*. Hillsdale, NJ: Erlbaum.

Moerk, E. L. (1980). Relationships between parental input frequencies and children's language acquisition: a reanalysis of Brown's data. *Journal of Child Language* **7**, 105–18.

Naigles, L. R. & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language* **25**, 95–120.

Ninio, A. (1996). Pathbreaking verbs in syntactic development. *Paper presented at the 7th International Congress for the Study of Child Language*, Istanbul, Turkey.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language* **26**, 619–53.

Okubo, A. (1967). *Yooji gengo no hattatsu* (*Children's language development*). Tokyo: Tokyodoo.

Peters, A. M. (2001). Filler syllables: what is their status in emerging grammar? *Journal of Child Language* **28**, 229–42.

Pine, J. M., Lieven, E. V. M. & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics* **36**(4), 807–30.

Pinker, S. (1981). On the acquisition of grammatical morphemes. *Journal of Child Language* **8**(2), 477–84.

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

Pinker, S. (1989). *Learnability and cognition: the acquisition of verb–arguments structure*. Cambridge, MA: Harvard University Press.

Redington, M. & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences* **1**(7), 273–81.

Rowland, C. F. & Pine, J. M. (2000). Subject-auxiliary inversion errors and wh-question acquisition: what children do know? *Journal of Child Language* **27**, 157–81.

Savic, S. (1975). Aspects of adult–child communication: the problem of question acquisition. *Journal of Child Language* **2**, 251–60.

Seidenberg, M. S. & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science* **23**(4), 569–88.

Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (eds), *Studies of child language development*. NY: Holt, Rinehart & Winston.

Theakston, A. L., Lieven, E. V. M., Pine, J. M. & Rowland, C. F. (2000). The role of performance limitations in the acquisition of 'mixed' verb–argument structure at stage 1. In M. Perkins & S. Howard (eds), *New directions in language development and disorders*. New York: Plenum Press.

Theakston, A. L., Lieven, E. V. M., Pine, J. M. & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language* **28**, 127–52.

Theakston, A. L., Lieven, E. V. M., Pine, J. M. & Rowland, C. F. (in submission). Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language*.

Tyack, D. & Ingram, D. (1977). Children's production and comprehension of questions. *Journal of Child Language* **4**, 211–28.

Valian, V. (1999). Input and language acquisition. In W. C. Ritchie & T. K. Bhatia (eds), *Handbook of child language acquisition*. San Diego: Academic Press.

Wode, H. (1975). Some stages in the acquisition of questions by monolingual children. *Word, Special Issue on Child Language*, 261–310.