

**Item-based vs. rule-based learning: the acquisition of word
order in wh-questions in English and Norwegian**

MARIT WESTERGAARD
University of Tromsø

ABSTRACT

This paper discusses an item-based and a rule-based model of language acquisition in relation to children's acquisition of word order in wh-questions in English and Norwegian. Constructivist work argues that English-speaking children are sensitive to wh-word + auxiliary combinations in the input, while generative work asserts that children acquire a rule of subject-auxiliary inversion or generalized verb second at an early stage. The paper questions both approaches, re-evaluates some previous work, and provides some further data, generally concluding that the acquisition of wh-questions must be the result of a rule-based process. Based on variation in adult grammars, an extended cue-based model to language acquisition is presented, according to which children are sensitive to so-called micro-cues in the input. Inversion is not considered to be one major rule, but several smaller-scale cues, which means that the full complexity of inversion may not necessarily fall into place at once.

INTRODUCTION

Children's acquisition of word order in non-subject *wh*-questions in English has been studied extensively, and in Ambridge, Rowland, Theakston & Tomasello (2006: 520), this is described as a topic which 'represents an ideal 'test case' for both movement-based generativist accounts and competing constructivist accounts of language acquisition.' The present paper takes a fresh look at some child data on word order in *wh*-questions in English and a dialect of Norwegian, at the same time comparing a generative and a constructivist approach. The Norwegian data discussed are mainly taken from Westergaard (2003), while the English data have been investigated in Rowland & Pine (2000). Within a generative framework, Westergaard (2003) argues for the early acquisition of a verb movement rule in Norwegian, generally based on target-consistent production of verb second (V2) as soon as the relevant constructions appear in the child data. Rowland & Pine (2000), on the other hand, take a constructivist approach to the English data and argue for a distributional learning mechanism that reproduces frequent combinations of *wh*-words and auxiliaries, e.g. *what + did*. Some children's failure to produce subject-auxiliary inversion is argued to be due to the target combinations being infrequent in the input.

The paper also provides some new data from the same children on word order in related constructions, e.g. embedded questions and questions with complex *wh*-phrases. I also present an extended cue-based model of language acquisition, arguing that children are sensitive to so-called micro-cues in the input. This may account for the attested micro-variation that is found across languages and dialects with respect to inversion, as well as children's general ability to detect minor distinctions in the input at an early age.

The paper is organized as follows: The next section provides an overview of the word order in *wh*-questions in English and the Tromsø dialect of Norwegian. I then briefly review the issue of item-based vs. rule-based learning and some previous accounts of the acquisition of word order in the relevant constructions. In the following section I outline the cue-based model and formulate the micro-cues that account for the word order variation. In the remainder of the paper I re-evaluate both the Norwegian and the English child data and conclude that it is difficult to analyze the word order of *wh*-questions without assuming some kind of rule-based process.

THE STRUCTURE OF THE TARGET LANGUAGES

Present-day English displays subject-auxiliary inversion in *wh*- and yes/no-questions, as illustrated in (1) and (2). In addition to affecting auxiliaries HAVE, BE and the modals, this process also applies to the copula BE. Subject-auxiliary inversion is a syntactic requirement, shown by the insertion of dummy DO when there is no other auxiliary present, as in (3).

- (1) What will Peter think?
- (2) Is Peter the man of your dreams?

- (3) What did Peter say? /*What said Peter?

This type of inversion is often called ‘residual V2’ (Rizzi 1996), as it is considered to be a remnant feature of a V2 grammar that existed in Old and Middle English. In present-day V2 languages, such as Norwegian, German or Icelandic, other clause types also display this word order, typically declaratives. In traditional generative accounts of V2 (e.g. den Besten 1977, Vikner 1995), it is assumed that the finite verb moves (from the Inflection position, I) across the subject to the head of the clausal projection (the Complementizer, C), thereby ending up in second position.

With respect to the word order of questions, the difference between English and Norwegian is that, while V2 is restricted to specific verb types in English (auxiliaries and BE), any lexical verb may appear in front of the subject in Norwegian, see (4). Strictly speaking, this is not a restriction on the V2 process *per se*, but a result of the fact that English has also lost V-to-I movement for lexical verbs (see e.g. Lightfoot 1999). And only what appears in the I position may move on to C in questions.

- (4) Hva sa Peter?
 what said Peter
 ‘What did Peter say?’

However, many dialects of Norwegian do not have a strict V2 requirement in *wh*-questions, allowing non-V2 in certain question types. Vangsnes (2006) gives an overview of the variation across dialects and argues that this is the result of several microparameters, mainly related to the type and function of the *wh*-element. The dialect under investigation (Tromsø) makes a distinction based on the length of the *wh*-constituent: While monosyllabic *wh*-elements allow both word orders, the longer ones (disyllabic and full *wh*-phrases) require V2, as illustrated in (5) and (6).

- (5) Ka les du? / Ka du les?
 what read you
 ‘What are you reading?’
- (6) Korfor kommer du? /*Korfor du kommer?
 why come you
 ‘Why are you coming?’

Furthermore, subject questions require non-V2 in the form of the relative complementizer *som* in second position, see (7). This means that a subject question without *som*, found in standard Norwegian, is ungrammatical in the dialect.

- (7) Kem som kommer? /*Kem kommer?
 who SOM come
 ‘Who is coming?’

The variation between V2 and non-V2 with the short *wh*-elements is not random. A study of adult spontaneous speech has shown that there are clear preferences related to subject and verbs types, non-V2 typically appearing when the subject is a personal pronoun, and V2 when the subject is a full DP

and the verb is BE, see Westergaard (2003, 2005). This indicates that information structure plays a role, non-V2 being chosen with informationally given subjects and V2 with new or focused subjects. Examples are provided in (8) and (9).

- (8) kor er mitt fly? (INV, file Ole.17)
 where is my plane
 'Where is my plane?'
 (9) kor vi lande henne? (INV, file Ole.17)
 where we land LOC
 'Where do we land?'

Finally, both in Norwegian and English, embedded questions are different from main clause questions in that there is no inversion, see (10) and (11). However, there are varieties of English which do allow inversion in these cases, e.g. Belfast English and Indian Vernacular English, illustrated in (12) and (13), (from Henry 1995: 106 and Bhatt 2004: 1020).

- (10) Jeg vet ikke [hva Peter leser] / *hva leser Peter.
 I know not what Peter reads / what reads Peter
 'I don't know what Peter is reading.'
 (11) Let me show you [what I am reading] / *what am I reading.
 (12) He didn't say why had they come.
 (13) I wonder where does he work.

From an acquisition perspective, this word order variation in adult grammars means that there is quite a bit of detail that must be learned from input: Children acquiring English must learn that inversion only applies in questions and not in e.g. declaratives, and furthermore, that it is restricted to certain classes of verbs (auxiliaries and BE). Norwegian children growing up in Tromsø must learn that the rule applies in questions with some wh-elements but not others, and that in 'optional' contexts (with the short wh-words), it is dependent on the type of subject (given vs. new information, typically manifested as pronouns vs. full DPs). And both learners of (standard) English and Norwegian must learn that there is no inversion in embedded questions.

In the next section I briefly discuss the main differences between a generative and a constructivist approach and present some previous accounts of the acquisition of word order in wh-questions.

ITEM-BASED VS. RULE-BASED LEARNING

Within the generative literature it is typically argued that Universal Grammar (UG) provides the language-learning child with the necessary functional structure and constraints, and that all the child needs to do is to learn lexical items and the setting of certain language-specific parameters, e.g. +/- head-final or +/-V2. One of the main reasons for postulating the existence of parameters is the ease and speed of language acquisition; children should only need to be exposed to a few relevant examples to set a parameter and thus make major generalizations for their language. Target-consistent production at a certain level (often 90%) is then simply taken to be evidence of

acquisition. A constructivist approach, on the other hand, argues that children's early multi-word utterances do not reflect the existence of a 'big rule', but are simply the result of a functionally-based distributional analysis of the input. According to Tomasello (2003), the relative frequency of forms in the primary linguistic data (PLD) should account for the order of acquisition as well as children's non-target-consistent utterances. A complex linguistic system WILL eventually emerge, but this is a process that takes time, and in children's early production, there is little or no syntactic structure.

The acquisition of V2 has been studied extensively, see e.g. Poeppel & Wexler (1993) for German, Jordens (1990) for Dutch, Santelmann (1995) and Platzack (1996) for Swedish, and Westergaard (2003) for Norwegian. Generally, V2 is attested as soon as the relevant constructions appear in the child data, and wh-questions seem to be attested particularly early with V2. Santelmann (1995) finds that, while there are occasional examples of non-target-consistent word order in declaratives in child Swedish, the children's wh-questions are virtually without mistakes. Investigating child German, Clahsen, Penke & Parodi (1993/94) also find that questions with the wh-words *was* 'what' or *wo* 'where' plus the copula are attested extremely early with V2, and these questions are therefore considered to be rote-learned as unanalyzed chunks at the initial stage.

English children's acquisition of word order in wh-questions has also been the focus of many studies, and findings vary somewhat; see Ambridge *et al.* (2006) for an overview. While some studies report that target-consistent word order falls into place relatively early (as soon as auxiliaries appear in the child data, see e.g. Ingram & Tyack 1979, Radford 1992), other studies find a certain delay in subject-auxiliary inversion. This delay is normally not random, but seems to affect only some wh-elements or some auxiliaries. Thus, generative accounts typically argue that subject-auxiliary inversion is in place early, and occasional non-target-consistent forms are explained as a result of children's problems with particular wh-items (e.g. adjuncts *how* and *why* vs. arguments *what* and *who*, see DeVilliers 1991) or problems with certain auxiliary types (dummy DO or copula BE vs. the rest, see e.g. Santelmann, Berk, Austin, Somashekar & Lust 2002). Within the framework of Role and Reference grammar, Van Valin (2002) claims that the delay affects certain modals and all negated auxiliaries, the reason being that young children do not interpret these as tensed elements.

A constructivist account of the acquisition of wh-questions in English is advocated by Rowland & Pine (2000, 2003) and Rowland, Pine, Lieven & Theakston (2003). This approach basically assumes that English-speaking children's early wh-questions are the result a distributional learning mechanism that reproduces lexically-specific combinations of a wh-word and an auxiliary, e.g. *what* + *will*. Rowland & Pine (2000), henceforth R&P, further argue that wh-questions that appear with correct subject-auxiliary inversion contain wh-word and auxiliary combinations that are frequent in the input, e.g. *where* + *does*, while non-target forms appear when children attempt to produce questions with wh-words and auxiliaries that happen to be infrequent

combinations, e.g. *why + can't*. R&P have investigated the data from Adam in the Brown corpus in the CHILDES database (Brown 1973, MacWhinney 2000) and compared this to a sample of the mother's data. They find that there is a correlation between Adam's wh-questions and the frequency of wh-word + auxiliary combinations in the input, and conclude that there is no evidence 'to support the claim that the child [is] operating with a subject-auxiliary inversion rule applied to grammatical categories' (p. 179).

The role of the input is somewhat modified in Rowland & Pine (2003: 211), where they state that they do not wish to claim 'that all children's acquisition data can be explained in terms of input frequencies.' Ambridge *et al.* (2006) report on an experimental study which replicates the R&P finding that children have problems with wh-word + auxiliary combinations rather than these elements individually, e.g. that *what + do* is more problematic than *what + does*. Thus, they also argue that children are sensitive to lexically-specific combinations and do not have categories such as nouns and verbs in their grammars. However, the children in their study turn out to have problems with different wh-word + auxiliary combinations than Adam, and Ambridge *et al.* (2006: 543) acknowledge that 'a simple input-frequency based account cannot predict the wh-word + auxiliary combinations with which children will produce inverted and non-inverted questions.'

In Ambridge *et al.* (2006) it is claimed that the studies discussed are specific to English and do not address the full range of wh-questions found in other languages. However, both the generative parameter-based approach and the constructivist item-based account make claims about the general learning process, and these should in principle be extendable to other languages and other syntactic phenomena. The Norwegian child data from Westergaard (2003) and the English data investigated in R&P are therefore re-evaluated below. Given the word order variation outlined in the previous section, a parameter-setting account would predict massive overgeneralization of V2 or subject-auxiliary inversion in early child data. This would also create problems with respect to negative evidence in the unlearning process of these overgeneralized forms. A constructivist account would initially fare somewhat better. However, also this approach would predict overgeneralized word order from highly frequent combinations to less frequent ones (e.g. from main to embedded clauses). In the next section I therefore outline a cue-based model of language acquisition, which is a generative model that accounts for word order variation in the input terms of so-called micro-cues.

A MODEL OF MICRO-CUES

In order to account for the word order variation across classical V2 languages and English with respect to verb movement, Westergaard (2007) has developed an extended version of Lightfoot's (1999, 2006) cue-based approach to language acquisition and change. In the original model, a cue is a piece of structure which is produced in children's I-language on exposure to triggers expressed by certain sentence types in the PLD. In Lightfoot (2006: 86) the cue for V2 syntax is formulated as in (14), i.e. a structure with a finite verb in

C, which is expressed in all non-subject-initial clauses. Furthermore, it is argued that there must be a UG requirement that the cue is obligatory; otherwise children acquiring a V2 grammar would be unable to generalize this word order to all its contexts.

(14) $_{CP}[XP \text{ } _C V \dots]$

However, Westergaard (2007) discusses word order variation across the Germanic languages and points out that it is not the case that the verb is always in C. Generally, there is no V2 in embedded contexts, except in Icelandic and Yiddish. Exclamatives also normally exhibit non-V2. And in this paper we have already seen some of the variation that exists in Norwegian and English. Since all these grammars are obviously learnable, children must be paying attention to finer linguistic distinctions than what the cue in (14) states. For example, children must be sensitive to different kinds of C (i.e. different clause types), different kinds of initial elements (e.g. long or short *wh*-words), or different classes of verbs (e.g. auxiliaries and/or BE vs. lexical verbs), see also Roeper (2007). Moreover, learners must also at some point distinguish between cases where there is optional V2 that is dependent on information structure.

This means that children must be sensitive to smaller-scale cues, what Westergaard (2007) calls micro-cues, see also Lightfoot & Westergaard (2007). First and foremost, learners of Norwegian and English must distinguish between declaratives and questions, since only the latter clause type displays verb movement in English, while there are more complex distinctions between the two in Norwegian dialects. This can be solved by having separate cues for declaratives and questions, which is possible in a Split-CP model such as Westergaard & Vangsnes (2005) or Westergaard (2007), where different clause types are argued to have different heads in the CP domain depending on illocutionary force, declaratives being Top(ic)Ps and *wh*-questions Int(errogative)Ps. Lightfoot's cue for V2 can therefore be reformulated as (15) and (16). While (15) is generally not expressed in the PLD that English children are exposed to, Norwegian children produce this structure in their I-language grammar as a result of the appropriate input.

(15) Cue for V2 in declaratives: $_{TopP}[XP \text{ } _{Top} V \dots]$

(16) Cue for V2 in *wh*-questions: $_{IntP}[(wh) \text{ } _{Int} V \dots]$

With respect to (16), matters are somewhat more complex. Since V2 in English is restricted to auxiliaries and BE, this needs to be part of the cue, which could be formulated as (17). This cue specifies that Inflectional elements (i.e. I) should be found in the Int° head. Some other cue (not relevant here) will then take care of which elements may appear in I in English.

(17) Cue for V2 in *wh*-questions (English): $_{IntP}[(wh) \text{ } _{Int} I \dots]$

The cue for V2 in Norwegian *wh*-questions must also be more specific. First and foremost, there must be a distinction between long and short *wh*-elements, which, according to Westergaard (2005), corresponds to a difference between phrases and heads. Thus, the cue in (18) specifies that the

verb is obligatorily in Int^o when the *wh*-element is phrasal. The cue in (19), on the other hand, contains a *wh*-element which has been reanalyzed as a head (due to an economy principle, Head Preference, see van Gelderen 2004), and which therefore may appear in the Int^o position itself. Westergaard then argues that the verb movement that sometimes applies in these cases is the result of a lower head in the CP domain (see Rizzi 1997, 2001), the LowTopP, which is sensitive to information structure and attracts the verb only when the subject is new or focused information, marked [+FOC] here.

(18) Cue for V2 in questions with long *wh*-elements: IntP[XP_[+wh] Int^o V...

(19) Cue for V2 in questions with monosyllabic *wh*-elements:
IntP[Int^o [*wh*] LowTopP[LowTop^o V XP_[+FOC]] ...

But how do we account for embedded *wh*-clauses, which lack verb movement in both languages? According to the Split-CP model adopted here, different illocutionary force is reflected as different heads in the CP domain. Embedded *wh*-clauses, lacking interrogative force, are thus considered to be WhPs, not IntPs. This means that the cues as formulated in (17)-(19) exclude embedded *wh*-clauses, and nothing needs to be said about them for the languages at hand, standard English and Tromsø Norwegian. For languages that do display verb movement in these clause types (e.g. Belfast English), a separate cue involving the WhP would have to be formulated.

Finally, I would like to emphasize that what is important about this model is that this is a generative approach that does not seek to explain children's acquisition by reference to one major parameter. Instead, when acquiring inversion in *wh*-questions, children must learn a set of smaller-scale rules, formulated as micro-cues. This approach also differs from a constructivist approach in that the cues are not surface strings of word combinations, but pieces of (hierarchical) I-language structure that make reference to linguistically relevant distinctions.

A RE-EVALUATION OF THE NORWEGIAN CHILD DATA

Introducing the data

The Norwegian child data studied in Westergaard (2003) are taken from a corpus collected at the University of Tromsø (Anderssen 2006); see Table 1 for an overview.

TABLE 1. *Overview of the Norwegian corpus of child language, Tromsø dialect*

Name of Child	Age	Files	Child Utterances
Ina	1;8.20-3;3.18	Ina.01-27	20,071
Ann	1;8.20-3;0.1	Ann.01-21	13,129
Ole	1;9.10-2;11.23	Ole.01-22	13,485
Total			46,685

Westergaard focuses on the acquisition of word order in questions with monosyllabic wh-words, where the target grammar allows both V2 and non-V2. The main finding is that both word orders are attested in the child data as soon as the relevant constructions appear. Table 2 contains an overview of all complete non-subject wh-questions produced before the age of three.

TABLE 2. *Non-subject wh-questions with V2 and non-V2 word order in Norwegian corpus of child language*

WH-WORD	INA.01-23, age 1;8.20-2;10.12		ANN.01-21, age 1;8.20-3;0.1		OLE.01-22, age 1;9.10-2;11.23		Total
	Non-V2	V2	Non-V2	V2	Non-V2	V2	
<i>ka</i>	89	48	53	19	0	1	210
‘what’	(65%)	(35%)	(73.6%)	(26.4%)			
<i>kor</i>	15	128	15	63	0	42	263
‘where’	(10.5%)	(89.5%)	(19%)	(81%)			
<i>kem</i>	8	21	3	9	0	3	44
‘who’	(27.5%)	(72.5%)	(25%)	(75%)			
Total	112	197	71	91	0	46	517
	(36%)	(64%)	(44%)	(56%)			

Moreover, the preference for subject and verb types with the two word orders is very similar to that of the adult data, as illustrated in (20) and (21), V2 typically appearing with full DP subjects and the verb BE, and non-V2 with pronominal subjects (cf. examples (8) and (9) above). Note that the child Ole produces only V2, but all his full wh-questions are of the kind that would require this word order also in the production of the other speakers. He also produces certain questions without the wh-word (see below), and some of these display inversion while others do not, the choice corresponding to the adult pattern. Based on this, Westergaard concludes that V2 is acquired early and that children have an early sensitivity to information structure.

- (20) *kor e babyen?* (Ina.06, age 2;1.0)
 where be.PRES baby.DEF
 ‘Where is the baby?’
- (21) *ka ho har der # nedi?* (Ina.02, age 1;10.4)
 what she have.PRES there down-in
 ‘What does she have in there?’

But target-consistent production does not in itself provide evidence that a rule is acquired. A constructivist account of the findings could argue that the Norwegian children are simply reproducing certain word combinations that are frequent in the input, e.g. *kor er* ‘where is’ or *ka har* ‘what has/have’. In the next section we consider such an approach to the Norwegian data.

Word combinations and frequencies

As mentioned in section 2, not just auxiliaries, but any lexical verb can appear in second position in Norwegian wh-questions. This means that the number of possible wh-word + verb combinations is considerably higher than

in English. This fact alone could make a constructivist approach to these child data less plausible. But that depends of course on how many combinations that are actually produced at an early stage. R&P find that during Adam's so-called uninversion period (age 2;11.19-3;8.14), he produces 26 different wh-word + auxiliary combinations with inversion, 20 that occur uninverted only, and 3 that appear with both word orders. For comparison, I consider the data from only one of the Norwegian children, Ina, who has been recorded slightly longer than the other two (until approximately age 3;3). In addition to 309 wh-questions produced with the mono-syllabic wh-words (cf. Table 2), Ina produces 187 further main clause non-subject wh-questions in the corpus, making the total 496. A study of these questions reveals that she produces 63 wh-word + verb combinations.

But we also need to consider Ina's questions with target-consistent non-V2. Under a constructivist approach, the wh-word would in these cases presumably be considered to form a linguistic collocation with the subject, e.g. *ka du* 'what you' or *kor han* 'where he'. In Ina's data there are 40 different wh-word + subject combinations, making the total number of attested initial two-word combinations 103 (each one attested on average 4.82 times, 496/103). This number is of course considerably higher than Adam's 26 target-consistent combinations, and also produced at an earlier age, but in itself, this number obviously does not prove anything.

Given the number of possible subjects, it seems somewhat odd to assume that a wh-word and a subject should be a collocation that children pay attention to. An alternative (constructivist) approach to these non-V2 cases could be that they are the result of a process where the wh-word by itself is stuck onto initial position, while the rest of the sentence remains as before (subject before verb). However, if this were the case, the choice of element(s) to place clause-initially would still have to be dependent on the type of subject. This would be necessary to ensure that wh-words appear clause-initially mainly when the subject is a pronoun and wh-word + verb combinations when the subject is a full DP. That is, the Norwegian children would need to be sensitive to longer combinations than the two first words. This is not in principle problematic for a constructivist account: In order to explain some of the data that do not conform to their expectations, R&P (p. 178) and Ambridge *et al.* (2006: 544) suggest that the children in their studies are sensitive to larger formulae in the input, e.g. *why don't you* or *what do you*. But for the Norwegian data this would have to hold for virtually ALL the complete (three-word) wh-questions produced, of which there are several hundred different ones attested in the data. To my mind, this seems more than unlikely.

A possible constructivist objection to this might be to say that Ina's grammar DOES consist of syntactic categories at this stage, but no movement rules. This would make it possible to argue that the grammar contains two schemas for wh-questions, something like $[WH+BE+DP]$, producing V2, and $[WH+pronoun+V]$, resulting in non-V2. However, like adults, the children also produce the respective word orders with other combinations of subject and verb types, dependent on the information structure of the question. Sentence (22), for example, which has non-V2 and a full DP subject, is

uttered in a situation where *løva* ‘the lion’ was mentioned in the immediately preceding context (thus given information), while example (23), which has V2 and a pronominal subject, is produced in a situation where the child is pointing, and thus the subject *han der* ‘he there’ is somehow focused.

- (22) ka [/] ka løva like å spise mamma? (Ann.15, age 2;6.21)
 what what lion.DEF like.PRES to eat mommie
 ‘Mommie, what does the lion like to eat?’
- (23) ka hete han der? (Ina.07, age 2;1.23)
 what is-called he there
 ‘What is HE called?’

Another possible objection would be that although Ina’s grammar seems to be rule-based when the entire period is investigated, this does not exclude the possibility that there is item-based learning taking place at an early stage. A study of the earliest files (Ina.01-10, up to age 2;3.12) reveals that she produces only 72 wh-questions, and as many as 43 (59.7%) are of the kind *kor + er* ‘where + is’, which means that there may be some rote-learned forms at this stage. Nevertheless, Ina also produces 13 initial two-word combinations during this time, which means an average of 5.54, not very different from the overall average (4.82, see above). Thus, despite the existence of some very frequent combinations, there is no indication that Ina’s grammar is fundamentally different from what it is at later stages.

Finally, the existence of both V2 and non-V2 in the adult grammar means that we cannot detect clear word order mistakes in the Norwegian child data, and it is not possible to investigate the frequency of input in relation to target and non-target forms, as was done for English. We therefore move on to an investigation of other question types.

Questions with long wh-elements

In this section we consider the Norwegian children’s questions with long wh-elements, which require V2 in the adult language. This means that, in order to be target-consistent, the child grammar must treat the syntax of long and short wh-elements differently. A study of the three children’s production reveals that these questions are much less frequent and also appear considerably later than questions with monosyllabic question words.

TABLE 3. *Word order in questions with long wh-elements in Norwegian corpus of child language*

File (Age) \ wh-word	<i>korsen</i> ‘how’		<i>korfor</i> ‘why’		wh-phrases	
	V2	Non-V2	V2	Non-V2	V2	Non-V2
Ina.13-27 (2;5.25-3;3.18)	1	0	71	2	2	1
Ann.17-21 (2;8.4-3;0.1)	1	0	0	1	0	0
Ole.16-22 (2;8.5-2;11.23)	0	0	22	0	0	0
Total	2	0	93	3	2	1

Table 3 shows that these question types appear predominantly with target-consistent V2 (97/101, 96%), with the exception of one example in Ann’s

files and three in Ina's. Ann's parents come from an area north of Tromsø where the dialect marginally allows non-V2 also with the longer wh-elements (see Westergaard 2005), and they occasionally produce this word order themselves. Ann's one example is therefore presumably a result of this input. Ina's three examples are more difficult to explain, but the fact that they are so rare in her data (3.9%, 3/77) makes it possible to conclude that Ina nevertheless has a default V2 grammar with long question words. These data thus indicate that the children's grammar is like the adult grammar in that there is a different syntax for different wh-elements.

Embedded questions

Embedded questions are generally much less frequent than main clause questions, also in the adult input (see e.g. Westergaard 2007, Westergaard & Bentzen 2007). As mentioned above, embedded questions disallow inversion in Norwegian, which means that children's production of word order in these clauses could reveal certain properties of their I-language grammar. If learners are setting a major word order parameter or are sensitive to a cue such as (14), one would expect them to initially produce structures where the verb is always in C, i.e. overgeneralization of V2. Similarly, if children are merely paying attention to frequent word combinations in the input, one might also expect them to produce inverted word order in embedded questions.

Investigating a large number of English-speaking children up to the age of 5;2, Diessel & Tomasello (2001) argue that children's earliest sentential complement utterances do not have any hierarchical structure. The so-called main clause part of these utterances typically consists of one of a very small number of similar verbs, e.g. 'think' or 'know', and these are argued to be linear constructions, where the matrix verb is a formula that is simply stuck onto the beginning of the sentence, which remains an unembedded structure.

An investigation of the Norwegian data initially reveals findings that are similar to what Diessel & Tomasello attested for English. The three children produce a total of 108 embedded questions in the corpus, the majority of them (91) attested in Ann's data, and only 13 and 4 in Ina's and Ole's data respectively. While Ann's first example is attested at age 2;2.19 and Ina's as early 1;11.22, the majority of these questions appear in the latter half of the recordings. In all these examples there are indeed only three matrix verbs attested, *se* 'look', *vite* 'know' and *vis* 'show', illustrated in (24)-(26). So perhaps these are simply linear constructions, with a linguistic formula clause-initially, e.g. *se her* in (24)?

(24) *se her* [ka Ina gjør]. (Ina.04, age 1;11.22)
 look.IMP here what Ina do.PRES
 'Look here what Ina is doing.'

(25) Ann *vet ikke* [kor han er henne]. (Ann.09, 2;2.19)
 Ann know.PRES not where he be.PRES LOC
 'Ann doesn't know where he is.'

- (26) skal æ vise # [korsen man trøkke på knappen]? (Ole.20, 2;10.15)
 shall I show how one push.PRES on button.DEF
 ‘Do you want me to show (you) how you push the button?’

However, except for one example in Ina’s last file (which is possibly a restart, see Westergaard & Bentzen 2007: 280-1), these embedded questions all appear with non-inverted word order (107/108, 99.1%). This clearly shows that the Norwegian children distinguish between main and embedded structures.

There is even a set of examples in Ann’s data where the matrix clause clearly IS some kind of linguistic formula. These are sentences produced for a very brief period (files Ann.18-19, age 2;8.24-2;9.17), where the matrix clause consists of the string *vet du ka* ‘do you know what?’, which also occurs frequently without an embedded clause in the same files.¹ The result of this is a curious pattern of double *wh*-words, indicating that *vet du ka* is treated as a chunk. The resulting construction must nevertheless have some hierarchical structure, causing uninverted word order in the following clause, see (27) and (28). Main clause word order is unattested in these examples.

- (27) [vet du ka] [ka æ kan gjøre no]? (Ann.19, 2;9.17)
 know.PRES you what what I can do now
 ‘Do you know what what I can do now?’
- (28) [vet du ka] [kor æ har elefanten]? (Ann.19, 2;9.17)
 know.PRES you what where I have.PRES elephant.DEF
 ‘Do you know what where I have the elephant?’

Subject questions and wh-less questions

Recall from the overview of word order that subject questions always appear with non-V2 in the Tromsø dialect, as there is a requirement for the relative complementizer *som* in second position. A parameter-setting account would again predict overgeneralization of V2 in this case, while a constructivist approach would expect children to reproduce the relatively frequent initial word combinations, *kem som* or *ka som* ‘who/what SOM’. Unfortunately, there are few subject questions in the children’s production (as in the adult data), 12 in Ina’s, 7 in Ann’s and only 1 in Ole’s data. While all of Ann’s questions are target-consistent (the first attested at age 2;6.0), Ole’s one example and four of Ina’s early ones appear without *som*, see (29). This element is even left out in a situation where the child is imitating the adult, as in (30).

- (29) ka skjedde? (Ina.10, 2;3.12)
 what happen.PAST
 ‘What happened?’
 Target form: ‘Ka som skjedde?’

¹ In these two files there are altogether 12 examples with double *wh*-words, 2 examples with only one *wh*-word, and 10 examples of *vet du ka?* occurring by itself.

- (30) nei og nei ka skjer der. (Ole.06, 2;1.5)
 no and no what happens there
 ‘Oh no, what is happening there!’
 [= imitating INV saying <nei og nei ka som skjer der>].

These examples show that children do not simply copy frequent word combinations in the input - instead their production seems to be the result of the state of their grammar at the particular stage. However, these examples do not provide evidence for a parameter-setting account either, as it could simply be the case that *som*, like other functional elements, is somewhat delayed.

Finally, there is one more clause type in the Norwegian child data that should be mentioned, which also indicates that the children are not just reproducing initial word combinations. There is a considerable number of wh-less questions (altogether 143) in the child data, especially at an early stage. In these questions, subject-verb word order is completely independent of the presence of the wh-element, as verb movement has taken place across the subject in (31) but not in (32). Note, however, that the word order in these examples does reflect the information structure patterns typically found in the adult language.

- (31) er doktoren? (Ole.02, 1;10.0)
 be.PRES doctor.DEF
 ‘(Where) is the doctor?’
 (32) den gjør der? (Ole.02, 1;10.0)
 that do.PRES there
 ‘(What) is that doing there?’

This re-evaluation of the Norwegian child data shows that neither a traditional parameter-setting account nor a constructivist item-based approach seems able to explain the data. I now move on to a discussion of the English child data.

A RE-EVALUATION OF THE ENGLISH CHILD DATA

Introducing the data

In this section I reconsider the English child data investigated in R&P, viz. the production of Adam from the Brown corpus during the so-called uninversion period, age 2;11.28-3;8.14. As mentioned above, R&P find that there is a statistically significant correlation between Adam’s uninversion errors and the frequency with which the mother uses certain wh-word and auxiliary combinations, and in Rowland & Pine (2003: 203) they claim that investigating more of the input gives even clearer results.

In their analysis of Adam’s data, R&P have only considered wh-questions involving auxiliaries and disregarded BE. However, I would argue that it is important to include the copula in order to get a complete picture of children’s behavior with respect to inversion. In my re-evaluation of Adam’s data, I therefore include his wh-questions with BE, and furthermore, three instances of main wh-questions introduced by *when*, as well as occasional questions with full wh-phrases such as *what color* or *what kind of car*. I have

also counted examples somewhat differently from R&P – as repetitions of the child’s own or an adult’s previous wh-question have not been excluded. The reason is that very often children do NOT imitate word for word what has been said, see e.g. (30) above. Such examples are an important indication that the child’s grammar is not completely target-like. When a child actually does repeat a sentence in an identical fashion, that is also an interesting fact, in my view. Thus, my numbers are somewhat higher than the ones found in R&P.

Table 4 provides an overview of Adam’s wh-questions at the three data points identified by R&P (files 19-24, 25-30 and 31-36), according to my calculations. Not unexpectedly, the figures for auxiliaries are similar to R&P’s findings. But the numbers for BE are very different from this: Wh-questions with missing BE are much less frequent than with missing auxiliaries, while target-consistent word order is correspondingly more frequent. Throughout this period there is a gradual decrease in the former and an increase in the latter, so that at data point 3, Adam’s production is almost completely target-consistent with the copula. Furthermore, there are hardly any non-inversion errors with BE throughout the investigated period.

TABLE 4. *Overview of wh-questions with missing verbs/auxiliaries, non-inverted and inverted word order in Adam.19-36 (age 2;11.28-3;8.14)*

Files	BE			AUX		
	Missing	Non-inv.	Inversion	Missing	Non-inv.	Inversion
19-24	60.6% (152)	2.0% (5)	37.4% (94)	87.6% (254)	2.4% (7)	10% (29)
25-30	30.4% (97)	2.5% (8)	67.1% (214)	77.2% (247)	15% (48)	7.8% (25)
31-36	14.2% (25)	2.3% (4)	83.5% (147)	57.0% (130)	9.2% (21)	33.8% (77)

These results are also displayed in Figure 1, which additionally provides an illustration of Adam’s total production of wh-questions. Note that at the end of the period, there are considerably more target-consistent forms than wh-questions with missing verbs (56% vs. 38%), and furthermore, that during this so-called uninversion period, there is no point at which Adam produces more uninverted than inverted forms.

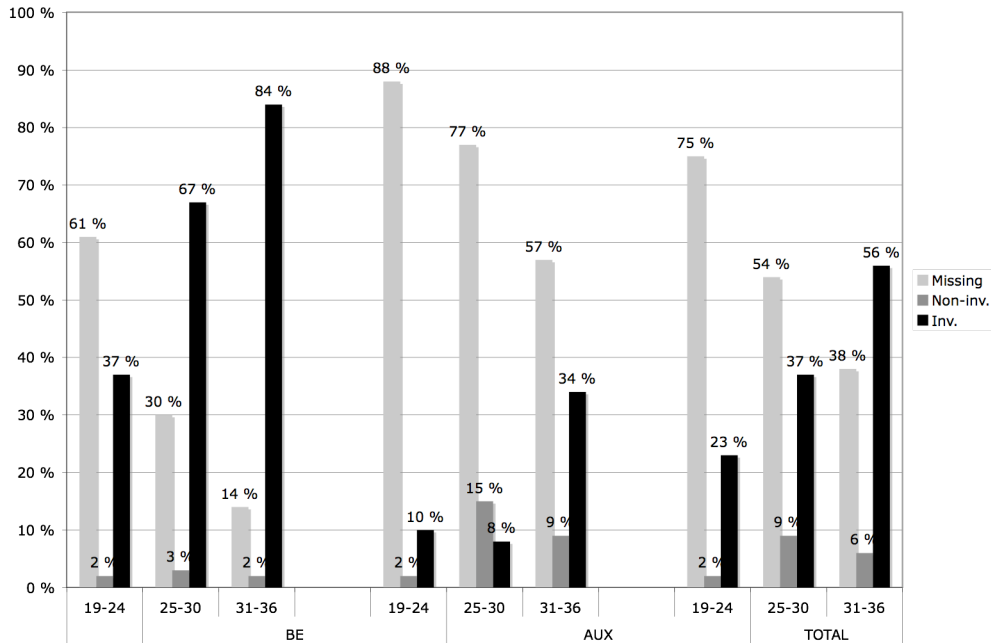


Fig 1. Percentage of wh-questions with missing verbs, non-inverted word order and inverted word order at three data points in Adam.19-36, age 2;11.28-3;8.14, with BE, AUX and the two categories combined.

Nevertheless, the important thing here is that Adam's grammar clearly makes a distinction between auxiliaries and BE. Inversion with BE does not seem to be a problem – whenever the copula is included it is virtually always inverted. The development with auxiliaries also seems to illustrate a learning path rather than a child gradually getting better at copying the input: During the first stage, Adam is producing a majority of verbless or auxiliary-less structures, and some of the target-consistent inverted structures are possibly rote-learned, as they cluster in some frozen formulae such as *how do you do it?* With respect to the non-target-consistent forms, we see typical U-shaped development. At data point 2, when there is a rise in uninversion errors, Adam may be realizing that his earlier unanalyzed chunks consist of separate words and he is experimenting with their order. Positive evidence in the input then steers him in the right direction, causing more target-consistent production at data point 3.

A comparison with the Norwegian data also seems relevant here. While the majority of Adam's wh-questions have missing verbs, the number of verbless wh-questions is relatively low in the Norwegian corpus, Ina producing 4.8% (22 out of 455 wh-questions – complete, verbless, and wh-less), Ann 3.0% (5/169), and Ole only 1.6% (1/64), all occurring in early files. In all cases the missing verb seems to be BE. This indicates that the high number of verbless questions in the English data is related to a problem with auxiliaries, or functional elements in general. Finally, let us also note that the figures for missing auxiliaries in Adam's data include a number of examples where the main verb is marked as finite (136/631, 21.6%), see (33) and (34).

- (33) what dey said? (Adam.24, age 3;2.09)
 (34) where firetruck goes? (Adam.20, age 3;0.11)

Questions with 'long' wh-phrases

In this section I consider questions with long wh-phrases in Adam's production. As in the Norwegian child data, such questions are much less frequent than those introduced by simple wh-words. Thus, if Adam is getting the word order right only in those questions that have wh-word + auxiliary combinations that are frequent in the input, as argued by R&P, questions with long wh-phrases should come out wrong most of the time. That is, combinations such as *what color + is* or *what kind of car + do* are predicted to appear uninverted. Adam's questions with long wh-words during the relevant period are displayed in Table 5.

TABLE 5. *Questions with long wh-phrases in Adam.19-36, age 2;11.28-3;8.14*

Files	Aux missing	Non-inversion		Inversion	
		BE	Aux	BE	Aux
19-24	0	1	0	4	0
25-30	3	1	2	25	0
31-36	1	0	0	3	3
Total	4	2	2	32	3

As many as 35 of Adam's 43 questions with long wh-phrases appear with target-consistent inversion (81.4%), illustrated in (35) and (36), only four are uninverted (9.3%), see (37), while four examples appear without an auxiliary. Again, there seems to be a difference between auxiliaries and BE: While BE is correctly inverted 94.1% (32/24), auxiliaries are inverted only 33.3% (3/9) uninverted 22.2% (2/9) and most often missing 44.4% (4/9).

- (35) What kind guns are dose? (Adam. 20, age 3;0.11)
 (36) What movie did I saw? (Adam.36, age 3;8.14)
 (37) Which bag they were # huh # in? (Adam.28, age 3;4.01)

These data clearly weaken R&P's claim that Adam's inversion pattern is not rule-governed, but only a result of frequent word combinations in the input. It should also be noted that in some of the target-consistent examples, the wh-phrase + auxiliary/BE combinations produced by Adam are so special that they have presumably never been heard before, e.g. (38) and (39).

- (38) Who glove is dis? (Adam.32, age 3;6.09)
 (39) What kind of monkey is he # playing da music? (Adam.30, age 3;5.01)

Embedded wh-clauses

Just like in Norwegian, there is no inversion in embedded questions in standard English. An investigation of the word order in this clause type may therefore again reveal important properties of the child grammar. Table 6

includes all Adam's embedded structures that are introduced by a wh-word; adverbial clauses as well as embedded questions.

TABLE 6. *Embedded wh-clauses in Adam.19-36, age 2;11.28-3;8.14*

Files	Non-inv. (Lex. verbs)	Non-inv. (Aux)	Non-inv. (BE)	Inv (BE)
19-24	2	0	2	1
25-30	38	5	7	2
31-36	35	3	5	3
Total	75 (72.8%)	8 (7.8%)	14 (13.6%)	6 (5.8%)

Most of Adam's embedded wh-clauses (75) contain lexical verbs, which do not invert with the subject in main clauses either. In all these cases, Adam produces target-consistent word order, as illustrated in (40), where the subject precedes the verb. In no case does he attempt to invert by adding dummy DO. Thus, these examples indicate that there is no 'transfer' of word order from main clause questions into embedded clauses.

(40) When you touch Paul # I spank you to pieces. (Adam.27, age 3;3.18)

More telling are examples with copula BE or auxiliaries, which SHOULD invert with the subject in main clauses. Twenty-two of the 28 examples appear without inversion, 14 with BE and 8 with auxiliaries, illustrated in (41) and (42). There are only 6 non-target examples, see (43), all with BE, again indicating that Adam's grammar distinguishes between the copula and auxiliaries. I return to a discussion of these examples below.

(41) So we can know where de mailman is. (Adam.25, age 3;2.21)

(42) I know what you can do. (Adam.34, age 3;7.07)

(43) I don't know what are they. (Adam.19, age 2;11.28)

This means that, just like the Norwegian children, Adam generally produces target-consistent non-inverted word order in embedded wh-clauses (97/103, 94.2%). This provides support for an analysis that assumes that his grammar has both structure and rules. A possible constructivist objection could be that Adam in these cases shows sensitivity to lexically-specific subject-verb combinations, which are of course also very frequent in the input. However, the production of non-target agreement in as many as 12 of the 22 examples with BE or auxiliaries makes such a hypothesis less plausible, see (44) and (45).

(44) Can I put my head in de mailbox # so de mailman can know where I are... (Adam.25, age 3;2.21)

(45) Once when you was showing me. (Adam 29, age 3;4.18)

To summarize so far, the new data on Adam's various wh-questions indicate that an item-based analysis such as the one advocated by R&P cannot account for his production. In the next section I also re-evaluate and question R&P's calculations of the relationship between Adam's data and the input.

Some further issues

As mentioned above, R&P find that Adam produces 26 combinations that are always target-consistent, 20 combinations that are always uninverted, and only 3 where both word orders are attested. The comparison of Adam's data to frequencies of wh-word + auxiliary combinations in a sample of the mother's data reveals that the 26 target-consistent combinations are attested 175 times in the input (an average of 6.73) while the 20 non-target-consistent combinations are only attested 35 times (average 1.75). In the larger input sample studied in Rowland & Pine (2003), there are 767 examples of the combinations that are always target-consistent in Adam's data (average 29.5), and only 247 examples of the always uninverted ones (average 12.5).

However, as R&P themselves acknowledge, as many as 21 of these combinations are only attested once in Adam's data (13 inverted, 8 uninverted); thus, there is no chance of overlapping word order in these cases.² Nevertheless, these are included in the comparison with the mother's data. R&P (p. 173) also emphasize that the six combinations that most frequently appear inverted in Adam's data (e.g. *what + do* 27 times, *what + are* 14 times) do not show any overlap. However, correct word order is in itself no evidence for the effect of the input, just like it cannot be evidence for the existence of a rule, as I argued above in connection with the Norwegian data. In my view, the interesting finding here is the 20 combinations that are always produced with non-target word order. Disregarding the 8 that only occur once, we are left with 12 combinations. But most of these are also quite rarely attested, generally only 2 or 3 times in the total data, and it therefore seems a bit bold to claim that they 'always' appear uninverted. In fact, there are only three non-target-consistent combinations which appear more than three times in the data, *what + can* (7), *why + can't* (10), and *why + don't* (6).

One also needs to consider the files before and after the uninversion period to be sure that Adam is consistently producing only one word order with these combinations. An investigation of Adam's wh-questions produced before the uninversion period reveals no conflicting examples. However, a study of his later wh-questions (files 37-55, age 3;8.26-5;2.12) shows that two of the combinations that are claimed to always be target-consistent in fact do appear with uninverted word order, see examples (46) and (47).

(46) where Paul's is ? (Adam.42, age 4;0.14)

(47) what Creepy Willy is ? (Adam.45, age 4;3.09)

A further problem is that as many as 12 of the 26 combinations that always appear with target-consistent word order involve dummy DO, e.g. *what + does* or *where + did*. But I would argue that these combinations are irrelevant as they are unlikely to appear uninverted. The dummy auxiliary in English is used to support either negation or a question feature. The latter will always appear in a high position (in C), while the former will occur in a low position (in I), below the subject. That means that there is no reason to insert

² For some reason, R&P (p. 173) claim that there are 19 combinations that only appear once, but a count in their own Table 3 (p. 172) reveals that there are in fact 21.

dummy DO in the low position unless there is negation; thus, children should not produce ungrammatical wh-questions such as *what she does like?* This also holds for Adam's data, as R&P have found only one such example, with the combination *why + did*. However, a closer examination of the context reveals that this is possibly not a true example of this combination, considering Adam's immediately preceding utterance, see (48). That is, Adam may be producing an auxiliary-less structure with a finite lexical verb, cf. (33) and (34) above, i.e. *why I did (it)*, attempting to say *why did I do (it) – break it?*

- (48) a. what was I did # break it?
 b. why I did break it? (Adam.28, age 3;4.01)

If the 12 wh-word + auxiliary combinations with dummy DO are excluded from R&P's list of formulae that are always produced with target-consistent word order, then the argument for the difference between the two lists with respect to input frequency is seriously weakened. The reason for this is that these 12 combinations constitute 80% of the input sample (140/175), while the remaining 14 combinations only account for 20%. This also holds for the larger input sample investigated in Rowland & Pine (2003), where the 12 combinations involving DO account for 80.2% (615/767).

As the importance of the input has been somewhat downplayed in more recent constructivist work, it may be that the authors would no longer claim that Adam's production is a direct result of the mother's input, and the sample studied in R&P is in any case clearly too small. As pointed out by Van Valin (2002: 164), as many as 11 of the 26 target-consistent and 15 of the 20 non-target-consistent combinations are not attested in the mother's input sample at all, and it is difficult to see how any prediction could be made based on this. In the larger sample of Rowland & Pine (2003), this still holds for 4 target-consistent and 10 non-target-consistent combinations. Furthermore, there is enormous variation within the lists, ranging from 1 to 169 examples in the input for the always target-consistent combinations *what + have* and *what + do* respectively. Similar numbers are found in the always uninverted pattern, from 1 example for *what + should* to 167 for *why + don't*. In this perspective, it seems odd to argue that *why + don't* is always uninverted in Adam's data because of lack of input, when it is fact attested much more frequently than 25 of the 26 always target-consistent combinations.

I would argue that the issues raised here would necessitate a different calculation of Adam's data in relation to the input sample. First of all, the 21 (13 + 8) combinations that are only attested once in Adam's data should be removed from the lists. The combinations that are not attested in the input at all should also be excluded (4 + 10 in Rowland & Pine 2003). And third, we need to exclude the 12 target-consistent combinations with dummy DO, as they arguably couldn't appear uninverted for independent reasons. If we now make a comparison between Adam's data and the input sample in Rowland & Pine (2003), the remaining lists would consist of 8 non-target-consistent combinations and only 2 target-consistent ones (*what + is*, *what + are*). But one of these was attested with uninverted word order in a later file, cf. (47),

and we are then left with only one always target-consistent combination. It should be obvious that a meaningful comparison could no longer be made.

AN ANALYSIS IN TERMS OF MICRO-CUES

In the previous sections I have re-evaluated both the Norwegian and the English child data and generally concluded that neither a parameter-setting nor a constructivist model such as the one advocated in R&P provides a satisfactory account of the findings.

Above I also briefly outlined a cue-based approach to language acquisition which is based on children discovering micro-cues in the input. These micro-cues reflect the fact that there is considerable variation across languages and dialects with respect to verb movement, which may be dependent on the type of wh-element, the class of verb, and the information value of the subject. This micro-variation must be learnable, which means that children at some point in the acquisition process must be sensitive to these distinctions. It is of course logically possible that children first use other learning strategies, e.g. set a major word order parameter or pay attention to frequent word combinations in the input, and then only later learn the fine distinctions and/or the deeper structure of the language they are exposed to. However, the child data considered above do not indicate that children go through such phases. Instead they seem to be sensitive to the micro-cues at an early stage. In the acquisition process, children must be conservative learners, not automatically (over-)generalizing inversion from one context to another. For example, Norwegian children immediately distinguish between long and short (phrasal and head-like) wh-elements as well as the information structure patterns relevant for verb movement. Most English-speaking children also master subject-auxiliary inversion in questions from early on and do not overgeneralize it to other clause types or other verb types (e.g. Radford 1992, Kuczaj & Maratsos 1983). And both Norwegian and English-speaking children distinguish between main and embedded contexts at an early stage.

Nevertheless, given the complexity of the variation it is not surprising that some children entertain slightly different verb movement grammars at the various stages of acquisition. That is, the full grammar of inversion may not fall into place immediately, some micro-cues lagging behind others. It is also not unlikely that individual children may acquire the micro-cues in different orders. This would then also explain the sometimes diverging findings across studies; different studies have investigated different children, and children may simply have different acquisitional paths.

Ambridge *et al.* (2006) argue against generative accounts, since these claim that children experience a delay either with respect to type of wh-word or type of auxiliary, while in their own study they find an effect of the combination of wh-words and auxiliaries. But according to the cue-based approach, it is not the case that a delay related to certain wh-words excludes the possibility of a delay with type of verb. After all, the variation across languages is not limited to one or the other, and it is clear that children learning the

different varieties must be sensitive to both types of distinctions. This also seems to be the case in Adam's development, which we now return to.

First of all, it is clear that Adam's grammar during the so-called uninversion period makes a distinction between auxiliaries and the copula (cf. Figure 1). That is, auxiliaries are more often missing, and inversion falls into place much later than with BE. Note also that in those cases where inversion is overgeneralized, in embedded clauses, this only affects BE (cf. Table 6). Early target-consistent word order with the copula cannot be the result of rote-learned forms only, as there are many examples in the data that do not belong to the typical formulae of frequent word combinations, e.g. the ones with long *wh*-phrases. The distinction between BE and auxiliaries is linguistically relevant and also found in other languages, e.g. in the Norwegian adult data, where inversion is clearly preferred with BE, but not auxiliaries. There is even evidence of this in English, as so-called stylistic inversion in declaratives mainly affects BE, but not auxiliaries (see e.g. Birner 1995). This means that Adam's grammar is at some point assuming even finer micro-cues than the target grammar.

As has been noted by R&P themselves and many others (e.g. Van Valin 2002), there is also a difference between *why* and other *wh*-words in Adam's production, noted also in data from other children. That is, Adam's grammar generally does not display inversion in *why*-questions (88.1%, 52/59); there are even uninverted examples with BE in these cases. In comparison, non-inversion with the question words *what* and *where* appear in only 3.4% (24/713) and 4.0% (7/173) of the data. According to Table 4, Adam produces a total of 76 uninverted examples with auxiliaries and 17 with BE. As many as 55.3% (42/76) of the former appear in questions with *why*, and 58.8% (10/17) of the latter. Many explanations for the special status of *why* have been explored, see e.g. Ambridge *et al.* (2006). I will not contribute to this discussion, but simply point out that this distinction is presumably also a linguistically relevant one, as *why* is syntactically different from other *wh*-elements also in other languages, e.g. Italian (see e.g. Poletto & Pollock 2004). Adam's behavior is thus not unlike that of Norwegian children, who have different V2 grammars for different *wh*-elements at an early age.

We then turn to the occasional examples in Adam's data of overgeneralization of inversion (with BE) in embedded clauses. Similar findings are attested in Swiss German child data (Schönenberger 2001) and in embedded declaratives in Norwegian (Westergaard & Bentzen 2007), argued in the latter to be due to an economy principle of movement. It is possible that Adam, like some other English-speaking children, has misinterpreted the cue for verb movement in main clauses to be only to a lower head in the CP domain, Wh° , and not Int° as in the adult grammar. Since embedded *wh*-clauses are considered to be bare *WhPs* in the Split-CP model (see above), this would mean that the grammar assumes that there IS inversion also in these embedded contexts. This is obviously a perfectly possible grammar, since both Belfast English and Indian Vernacular English display this word order. This tendency for economy will be overridden by input in standard English, and Adam apparently quickly reformulates the micro-cue to the

adult version, as there are very few non-target-consistent examples in his data.

Finally, it has also been noted that negation seems to play a role in English children's early *wh*-questions, in the sense that negated auxiliaries generally do not invert, see Ambridge *et al.* (2006) and references cited there. In R&P's list of *wh*-word + auxiliary combinations that consistently appear with uninverted word order in Adam's data, 7 of the 20 combinations contain a negated auxiliary (3 with DO, 4 with modals), making up 43.3% of the total (26/60). Of the 26 combinations that consistently appear with inverted word order, on the other hand, none are negated. The discussion of the role of dummy DO in Adam's grammar in the previous section may now shed some new light on why uninversion errors are relatively frequent with negated forms. Recall that I argued that there is no need to insert the dummy auxiliary unless it moves to the high position (to C) in questions. However, there IS a reason to insert DO in the low position when there is negation that needs to be supported. This could possibly also be used to account for the high frequency of negated modals in the non-target-consistent forms, as the modals may also have been inserted mainly to support negation. In a positive question, on the other hand, the modals may more likely simply be missing, as *wh*-questions with missing auxiliaries are still the most frequent question type in Adam's data, even at the end of the investigated period.

To conclude, I would argue that when children's utterances are non-target-consistent, this indicates that something is late acquired because it represents a functional element, or that the child grammar is different from that of the adult language in the sense that not all micro-cues are in place. Auxiliaries, being functional elements, should thus be missing more often than lexical verbs, a claim which is supported by a comparison of the English and the Norwegian child data. The functional element *som* in Norwegian subject questions may also be delayed for the same reason. Like any theory of language acquisition, also this one must allow for the existence of some rote-learning at an early stage. But when non-target-consistent production shows a principled pattern of e.g. uninversion, this indicates that children have only part of the inversion rule in place. In the process of discovering all the relevant small-scale cues, input frequency may play a role for the timing of acquisition or the persistence of misinterpretation, but, as also argued in Westergaard & Bentzen (2007), frequency is not by itself a cause for non-target behavior.

SUMMARY AND CONCLUSION

In this paper, I have discussed the issue of item-based vs. rule-based learning in relation to children's acquisition of word order in *wh*-questions in English and Norwegian. In my re-evaluation of previous work, I have questioned both a generative parameter-setting approach as well as a constructivist item-based approach, and based on some further data on different types of *wh*-questions, I have concluded that neither can account for the Norwegian and English child data. The main finding of the investigation of the Norwegian

data is that the children are perfectly able to make fine distinctions with respect to inversion at an early age, e.g. between long and short wh-elements or between main and embedded contexts. Furthermore, they do not seem to be paying attention to frequent word combinations in the input, both because of the high number of possible (and attested) combinations, as well as the fact that such combinations are not always imitated (in subject questions). A re-evaluation of the English data in R&P reveals that Adam also produces target-consistent word order in embedded questions and questions with long wh-elements. However, there are indications that inversion in Adam's grammar is only partly in place, as he makes a distinction with respect to type of wh-element (*why* vs. the rest) and type of verb (BE vs. auxiliaries). I also dispute R&P's calculation of Adam's production in relation to the input.

The general conclusion is that it is difficult to analyze the Norwegian and English child data on word order in wh-questions without arguing that the children's production is the result of some rule-based process. More specifically, I have presented a cue-based approach to language acquisition which asserts that children must discover minor (linguistically relevant) distinctions in the input, so-called micro-cues. In the case of inversion, these reflect the fact that there is considerable variation across adult languages and dialects with respect to the elements involved; wh-element, verb and subject. The child data investigated also indicate that children are sensitive to these micro-cues from early on, as they are generally able to make the fine distinctions immediately. Nevertheless, due to the complexity of inversion, some of the micro-cues may be delayed or misinterpreted, resulting in non-target-consistent, but still principled, behavior for a certain period of time.

REFERENCES

- Ambridge, B., Rowland, C. F., Theakston A. L. & Tomasello, M. (2006). Comparing different accounts of inversion errors in children's non-subject wh-questions: 'What experimental data can tell us?' *Journal of Child Language* **33**, 519-557.
- Anderssen, M. (2006). *The Acquisition of Compositional Definiteness in Norwegian*. Doctoral Dissertation, University of Tromsø.
- Bhatt, R. M. (2004). Indian English: syntax. In B. Kortmann, K. Burridge, R. Mesthrie, E. W. Schneider & C. Upton (eds.), *Handbook of Varieties of English 2: Morphology and Syntax*. Berlin/New York: Mouton de Gruyter. 1016-1030.
- Birner, B. (1995). Pragmatic constraints on the verb in English inversion. *Lingua* **97**, 233-256.
- Brown, R. (1973). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- Clahsen, H., Penke, M. & Parodi, T. (1993/94). Functional categories in early child German. *Language Acquisition* **3**, 395-429.
- den Besten, H. (1977). On the interaction of root transformations and lexical deletive rules. Ms., University of Amsterdam. Published 1981 in *Groninger Arbeiten zur Germanistischen Linguistik* **20**, 1-78; republished 1983 in W. Abraham (ed.), *On the formal syntax of the Westgermania*. Amsterdam: Benjamins, 47-131,

- republished 1989 in H. den Besten, *Studies in Germanic Syntax*. Amsterdam: Rodopi, 14-100.
- DeVilliers, J. (1991). Why question? In T. L. Maxfield & B. Plunkett (eds.), *Papers in the acquisition of wh: Proceedings of the Umass Roundtable, May 1990*. Amherst, MA: University of Massachusetts Occasional Papers.
- Diessel, H. & Tomasello, M. (2001). The acquisition of finite complement clauses in English: a usage-based approach to the development of grammatical constructions. *Cognitive Linguistics* **12**, 97-141.
- Henry, A. (1995). *Belfast English and Standard English*, New York and Oxford: Oxford University Press.
- Ingram, D. & Tyack, D. (1979). Inversion of subject NP and Aux in children's questions. *Journal of Psycholinguistic Research* **8.4**, 333-341.
- Jordens, P. (1990). The acquisition of verb placement in Dutch and German. *Linguistics* **28**, 1407-1448.
- Kuczaj, S. & Maratsos, M. (1983). Initial verbs of yes-no questions: a different kind of general grammatical category. *Developmental Psychology* **19.3**, 440-444.
- Lightfoot, D. 1999. *The Development of Language: Acquisition, Change and Evolution*. Malden, MA and Oxford: Blackwell.
- Lightfoot, D. (2006). *How New Languages Emerge*. Cambridge, UK: Cambridge University Press.
- Lightfoot, D. & Westergaard, M. (2007). Language Acquisition and Language Change: Inter-relationships. *Language and Linguistics Compass*.
- MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. 3rd Edition. Vol. 2: The Database. Mahwah, NJ: Lawrence Erlbaum Associates.
- Platzack, C. (1996). The initial hypothesis of syntax: a minimalist perspective on language acquisition and attrition. In H. Clahsen (ed.), *Generative Perspectives on Language Acquisition: Empirical Findings, Theoretical Considerations, Crosslinguistic Comparisons* [Language Acquisition and Language Disorders **14**], 369-414. Amsterdam/Philadelphia: John Benjamins.
- Poeppl, D. & Wexler, K. (1993). The full competence hypothesis of clause structure in Early German. *Language* **69**, 1-33.
- Poletto, C. & Pollock, J.-Y. (2004). On wh-clitics and wh-doubling in French and some North Eastern Italian dialects. *Probus* **16**, 241-272.
- Radford, A. (1992). The acquisition of the morphosyntax of finite verbs in English. In J. M. Meisel (ed.), *The Acquisition of Verb Placement: Functional Categories and V2 Phenomena in Language Acquisition*, 23-62. Dordrecht: Kluwer.
- Rizzi, L. (1996). Residual verb second and the wh-criterion. In A. Belletti & L. Rizzi (eds.), *Parameters and Functional Heads*. 63-90. Oxford: Oxford University Press.
- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (ed.), *Elements of Grammar: Handbook of Generative Syntax*, 281-337. Dordrecht: Kluwer.
- Rizzi, L. 2001. On the position 'Int(errogative)' in the left periphery of the clause. In G. Cinque & G. Salvi (eds.), *Current Studies in Italian Syntax*. 287-296. Amsterdam: Elsevier.
- Roeper, T. (2007). What frequency can do and what it can't. In I. Gülzow & N. Gagarina (eds.), *Frequency effects in language acquisition: Defining the limits of frequency as an explanatory concept*, 23-48. [Studies on Language Acquisition], Berlin: Mouton de Gruyter.

- Rowland, C. F. & Pine, J. M. (2000). Subject-auxiliary inversion errors and wh-question acquisition: 'What children do know?' *Journal of Child Language* **27**, 157-181.
- Rowland, C. F. & Pine, J. M. (2003). The development of inversion in wh-questions: a reply to Van Valin. *Journal of Child Language* **30**, 197-212.
- Rowland, C. F., Pine, J. M., Lieven, E. M. V. & Theakston, A. L. (2003). Determinants of acquisition order in wh-questions: re-evaluating the role of caregiver speech. *Journal of Child Language* **30**, 609-635.
- Santelmann, L. (1995). *The Acquisition of Verb Second Grammar in Child Swedish: Continuity of Universal Grammar in WH-questions, Topicalizations and Verb Raising*. Ph.D. dissertation, Cornell University.
- Santelmann, L., Berk, S., Austin, J., Somashekar, S. & Lust, B. (2002). Continuity and development in the acquisition of inversion in yes/no-questions: dissociating movement and inflection. *Journal of Child Language* **29**, 813-42.
- Schönenberger, M. (2001). *Embedded V-to-C in Child Grammar: The Acquisition of Verb Placement in Swiss German*. Studies in Theoretical Psycholinguistics. Dordrecht: Kluwer.
- Tomasello, M. (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Van Valin, R. D. (2002). The development of subject-auxiliary inversion in English wh-questions: an alternative analysis. *Journal of Child Language* **29**, 161-175.
- Vangsnes, Ø. A. (2006). Microparameters for Norwegian wh-grammars. *Linguistic Variation Yearbook* **5**, 187-226. Amsterdam and Philadelphia: John Benjamins.
- Vikner, S. (1995). *Verb Movement and Expletive Subjects in the Germanic Languages*. New York, Oxford: Oxford University Press.
- Westergaard, M. R. (2003). Word order in wh-questions in a North Norwegian dialect: some evidence from an acquisition study. *Nordic Journal of Linguistics* **26.1**, 81-109.
- Westergaard, M. R. (2005). Optional word order in wh-questions in two Norwegian dialects: a diachronic analysis of synchronic variation. *Nordic Journal of Linguistics* **28.2**, 269-296.
- Westergaard, M. R. (2007). Learning and unlearning V2: on the robustness of the triggering experience in a historical perspective. In M. Cecilia Picchi & Alan Pona (eds.) *Proceedings of the XXXII Incontro di Grammatica Generativa*, 193-207. Alessandria: Edizioni dell'Orso.
- Westergaard, M. R. & Vangsnes, Ø. A. (2005). Wh-questions, V2, and the left periphery of three Norwegian dialects. *Journal of Comparative Germanic Linguistics* **8**, 117-158.
- Westergaard, M. & Bentzen, K. (2007). The (Non-) effect of input frequency on the acquisition of word order in Norwegian embedded clauses. In I. Gülzow & N. Gagarina (eds.), *Frequency effects in language acquisition: Defining the limits of frequency as an explanatory concept*, 271-306. [Studies on Language Acquisition], Berlin: Mouton de Gruyter.