HS First Language Acquisition
SS 2004
Friedrich Schiller Universität Jena
Institut für Anglistik/Amerikanistik
P.D. Dr. Holger Diessel

The Acquistition of Wh-Questions: The Formula-to-Schema Hypothesis

Susann Keßler
Matr. Nr.: 33672
10.Semester

Lutherstr. 166
07743 Jena
e-mail: susann-kessler@web.de

Content

1. Introduction

When looking at children's early speech, we immediately realize that a large part of their utterances are questions. But when do children start to produce them and how do they actually form interrogative structures and what do they do when they learn them? How does a child learn to form an adult-like question such as:

a) What is Daddy doing in the garden at noon?
b) Why did Peter eat all the food at the party yesterday?

A) and b) are examples of adult-like wh-questions, consisting of a wh-question word, an inverted auxiliary, a subject and verb. Example a) contains an adverbial of place and time and example b) a direct object as well as adverbials of place and time.

The traditional answer to the question what children learn when they learn to form wh-questions is that children acquire some mental analogue of abstract rules like the wh-movement. These abstract rules are innate in our brains in the form of a Universal Grammar (UG), Chomsky and the nativists claim. There is, however, little positive evidence for children learning such an abstract transformational rule.

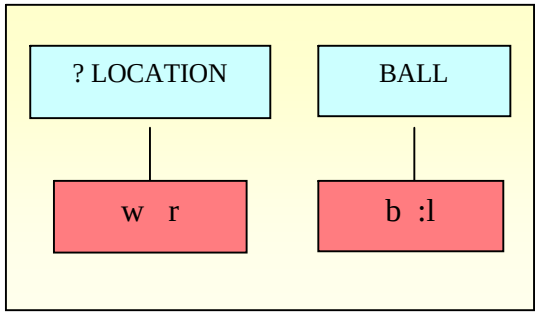
In November 2000, Ewa Dobrowsky, in her study of interrogative structures From Formulas to Schema: The Acquisition of English Questions posted the formula-to-schema hypothesis. She claims that there is evidence that children's early usage of questions is highly stereotypical and that "development proceeds in a piecemeal fashion"[1]. Thus, children start out with a fixed repertoire of "lexically based patterns" or so called "formulas".

Formulas are invariant and can be formulas such as *Whatssis?* or *Whothat?* or formulaic frames such as *What_____doing?*, *Where_____going?*. In other words, formulas are "big words", which consist of a phonological form and a semantic representation. They are symbolic units or so called form-meaning pairings. Formulas are learnt by rote and retrieved from memory when needed. Furthermore, data shows that children rely on these formulas and formulaic

[1] Ewa Dobrowska. „From Formulas to Schema: The Acquisition of English Questions." Cognitive Linguistics 11: 2000, 83-102. 83. All page references used within the text refer to this article.

frames to a very large extend and much longer than believed. Even when children generalize, Dowbrowska suggests, they do not rely on abstract innate rules but low-level templates. (84) The following example shows such a form-meaning pairing or formula for the simple question *Where ball?*. (Words in CAPITALS will be used as abbreviations for an expression's semantic structure and the phonological form will be represented in phonetic transcription.)

Figure 1



At this point the question arises, how can children develop syntactic rules out of these form-meaning parings? Chomsky and the nativists say that we do not extract rules from the input we are given but from our UG. In contrast to Chomsky's position, Dobrowska, who takes an approach from cognitive grammar, says that we acquire our syntactic rules out of the input we are given. We add new formulas in a piecemeal fashion, we analyse the rote-learnt formulas and gradually extract constructional schema from them.

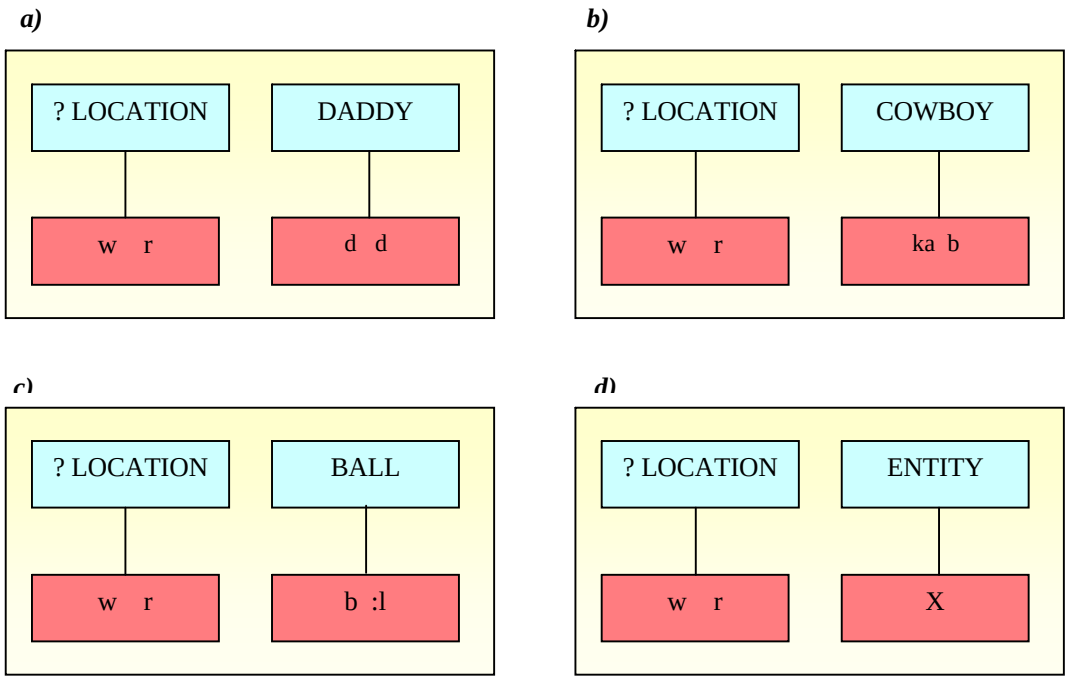
But how does the process of formula analysis and schema extraction work? The first phase, analysis, involves three steps:

1. The child segments the phonological representation. It segments the speech stream and finds out the different words of the utterance.
2. The child analyses the meaning. Using figure 1 as an example, /b :l/ refers to ball and /w r b :l/ means "What is the location of the ball?"
3. The child maps phonological material onto semantic structure. It now realizes that /w r/ is a request to give information about an entity's location.

Therefore, the constructional schema has the same structure as the formula preceding it as well as all the information which is implicit in the analysed

formula. Thus, an analysed formula is in fact a mini-grammar, a grammar for assembling one particular expression. To be able to construct novel utterance, the child needs a more general grammar – a constructional schema. (94) But how do children get from formula to schema? This happens in a second phase – extraction - which can be pictured as kind of "overwriting". The child, in this case Adam, acquires more and more formulas and the representations of the newly learnt ones overwrite the previous ones and finally result in generalization, a so called low-level schema. Figure 2 shows the development of such a slow-level schema.

Figure 2. ***Three formulas (a, b, c) and a low level schema implicit in them (d)***



The three formulas, a), b), and c) are superimposed on each other. The result would be a schema such as d). The shared parts [? LOCATION/ w r] remain unchanged, while the non shared parts, *Daddy, cowboy* and *ball*, are overwritten and generalized into [ENTITY/X]. At this point Adam has produced an abstract constructional schema and is now able to produce new, unheard questions such as, *Where book?* or *Where paperclip?*

Thus, a constructional schema can be defined as a symbolic unit which is complex and schematic, in other words more abstract and specified in less detail. It is the blueprint for assembling complex expressions and it is derived

from actual expressions and has the same structure as the expression it is derived from. (84/5)

Dobrowska argues that the starting point of syntactic development is a set of complex lexical units or formulas. Some formulas may remain "unopened" for some time but are eventually opened and schema are extracted. The schema extraction process is gradual, which means that the children progress "from invariant formulas through increasingly general formulaic frames to a constructional schema in which none of the slots are tied to specific lexical items" (87). To investigate this process of schema extraction a fairly large corpus of child utterances is necessary. The most suitable method is a case study. Dobrowska investigated one single child, Naomi, from the age of 1,6-3,8 years and posed the question whether "the approach is also applicable to other children." (84) It is obvious that further research is necessary to draw firm conclusions whether all children extract schema from formulas and acquire syntactic knowledge that way.

In my research, I took up Dobrowska's question and decided to investigate another child, Adam, with the same method in order to see whether Dobrowska's claims and the formula-to-schema hypothesis can be generalized.

2. Data

The data examined in my study were collected by Brown from 1962-1964, who donated them to the CHILDES database. It comprises 55 transcripts of Adam's speech from 2,3.4 (two years, three months and four days) to 4,10.23. Each transcript contains the data of 2 hours of recording, thus 110 hours total. The transcripts analysed for this study span the ages from 2,3.4 to 3,3.18, a total of about one year. Unfortunately, I had to limit the amount of data I analysed to a period of one year, because of the time limit of one semester for my research.

In a first step, I searched for all utterances containing a wh-word with the CLAN program. Later, all wh-questions were extracted and counted manually. In a next step, they were categorized into different question words such as

when, why where, who, what, etc.. In a final step, I determined which proportion of Adam's wh-questions are formulaic. In order to do that, all of his questions were categorized as different formulas, distinguishing between main formulas, minor formulas and non formulaic questions. For my research, I have adopted Dobrowska's definition of a formulas "which exploits the fact that combinations of words which recur again and again are likely to be stored, and regard as formulaic any sequence of simple units, with or without a slot, which occurs at least five times in" Adam's "corpus". (88) Sequences of units which are used at least ten times will be referred to as "major formulas". If a sequence occurs no more than four times in the corpus, the utterances have been classified as non-formulaic.

Some examples of Adam's formulas are given in example (1). The invariant parts of each formula are in *italics*, while CAPITALS represent slots which can be filled with simple words or phrases.

(1) *What* PERSON *doing*?
Where ENTITY *go*?
Which ENTITY?
What that got THING?
Who PROCESS *that*?
What kind THING that?
Why you PROCESS?

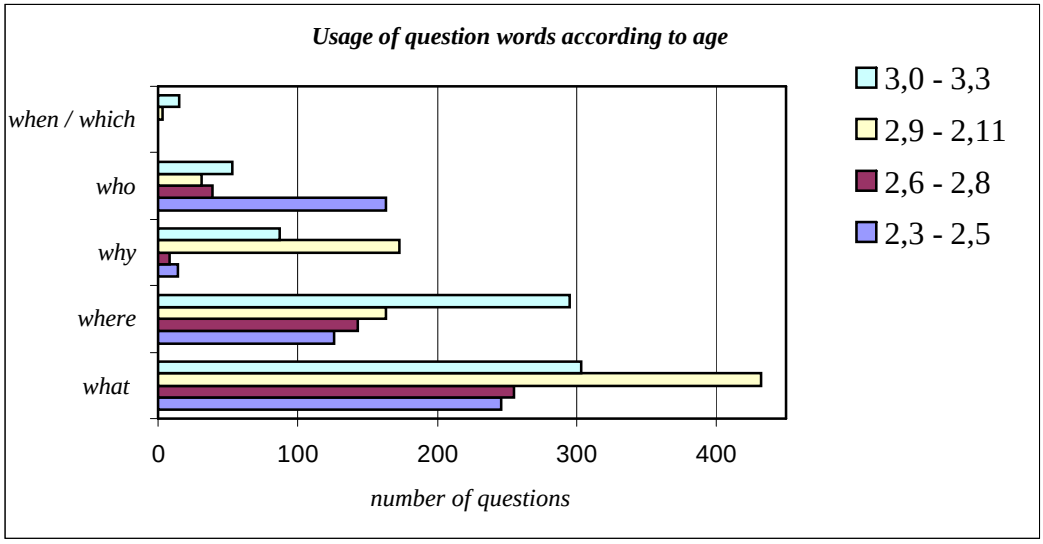
Some basic statistics about the total number of wh-questions formed by Adam from 2,3 –3,3 are given in table 1. Figure 3 shows the development of the different question words used by Adam in a graphic way.

Table 1. ***Total number of questions and question words***

age	what	where	why	who	when / which	total
2,3 - 2,5	246	126	14	163	0	**549**
2,6 - 2,8	255	143	8	39	0	**445**
2,9 - 2,11	432	163	173	31	3	**802**
3,0 - 3,3	303	295	87	53	15	**753**
total	**1236**	**727**	**282**	**286**	**18**	

F

Figure 3



Adam uses the wh-question words *what*, *where*, *why*, *who*, *when* and *which* from the age of 2,3-3,3. The wh-question word, *whose*, is absent from his speech. *When* is used twice at the age of 2, 11.28 (*When they driving?*) and 3,2.9 (*When this go on?*), whereas *which* is used 13 times in the analysed period. Out of the 13 times, the minor formula, *Which one?* occurs seven times. What-questions occur most frequently, over 1200 times, in Adam's speech, followed by where-questions, which he uses over 700 times. The frequency of why- and who-questions is about the same. Both are used over 280 times. The total number of Adam's questions increases. Figures 4 and 5 illustrate these findings.

Figure 4. ***Total number of question words***

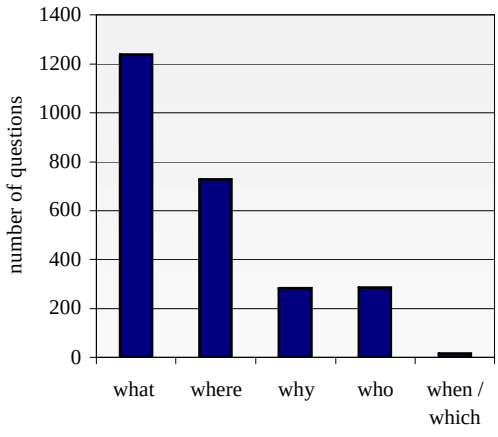
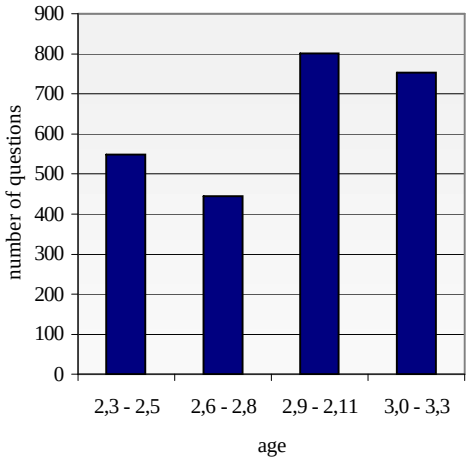


Figure 5. ***Total number of questions***



When looking at Table 1 and its graphic representation in figure 3, two numbers are striking. From the age of 2,3-3,5, Adam forms 163 who-questions. This number, however, drops rapidly and then slowly increases again as he gets older. A similar finding is the number of why-questions he uses between the age of 2,9 and 2,11. There seems to be sudden explosion in his development only to drop again rapidly during the next period of months. I will return to these issues later and try to give a sufficient explanation. For the moment, I want to come back to the formula-to schema-hypothesis and look at the formulaicity of Adam's questions which is shown in table 2. Figures 6 and 7 illustrate the results presented in table 2 in a graphic way.

Table 2 Formulaicity in wh-questions

age	major formulas	minor formulas	non-formulaic	total questions	% formulaic
2,3-2,5	511	14	24	549	95,60%
2,6-2,8	394	19	32	445	92,80%
2,9-2,11	660	36	106	802	86,80%
3,0-3,3	489	54	110	653	83,20%
total: 2,3-3,3	**2054**	**123**	**272**	**2449**	**88,89%**

Figure 6. ***Percentage of formulaicity***

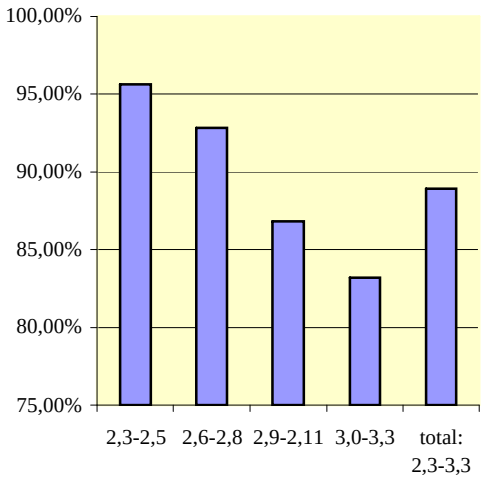
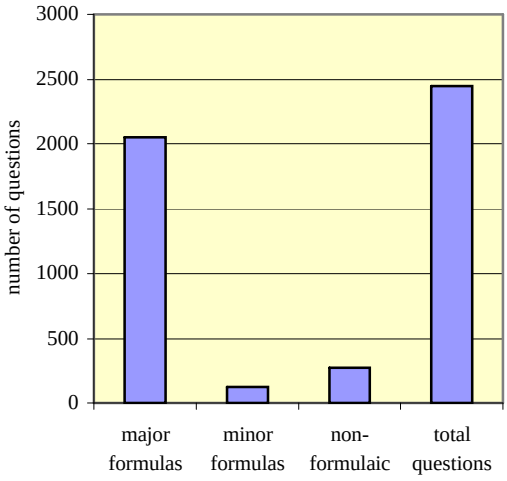


Figure 7. ***Distribution of formulaicity***



As we can see by the criteria defined earlier, the majority of Adam's wh-questions are formulaic. The overall proportion of formulaic questions in the analysed period is over 88 percent. Out of these 88 percent over 80 percent are major formulas. Comparing the numbers in the last column of Table 2, we see that the frequency of formulaic questions is highest in the earliest recordings (see Figure 7) and constantly drops as Adam grows older, whereas the total number of Adam's questions rises (see Table 1 and Figure 5). Furthermore, the percentage of Adam's non-formulaic questions also rises with his age. It should be pointed out, however, that these findings do not mean that the formula-to-schema hypothesis does not hold. Adam's questions, which occur no more than four times in his speech, are not necessarily non-formulaic. Formulas, like words, differ in frequency and are situation and context dependent. Adam's speech was only recorded every two weeks for two hours and therefore the questions classified as non-formulaic could be formulaic as well; Adam just does not use the formulas often enough because of the time limit and limit of situations and contexts, which is due to the recording periods. Therefore, we can be reasonably certain that a large number of Adam's non-formulaic questions are in fact memorized phrases. Furthermore, it is also obvious that Adam's early usage of wh-questions is highly stereotypical and becomes less stereotypical as he grows older. (89)

3. Analysis

The development of wh-questions in Adam's speech is somewhat paradoxical. The early recordings contain only 5 types of wh-questions. The invariant formulas: *What that?*, and *Who that?*, and the formulaic frames *Where* _____? (*Where Daddy?, Where doggie?, Where circus?, etc.*) and *Where* _____ *go*? (*Where tractor go?, Where Daddy go?, Where glove go?, etc.*) Therefore, it is clear that at the beginning of the period covered by this study, Adam's ability to form wh-questions is very limited. Adam's extensive usage (over 150 times) of the *Who that*-formula also explains the question about the high number of who-questions at the age of 2,3-2,5, which was raised

earlier. His repertoire consists of only one formula which he uses over and over again, in a very stereotypical way. However, it is equally clear that towards the end of the period at the age of 3,3, Adam commands a wide repertoire of wh-questions, which is shown by the following examples:

(2)	*Why he going to have some seeds?*	(3,2.21)
	What you going to give me?	(3,2.11)
	Where the sponge come from?	(3,3.4)
	Why you heard a little click?	(3,3.4)
	What is this record about?	(3,3.4)
	What is that man doing?	(3,3.4)
	What you doing with all those things in there?	(3,3.18)
	Who is coming?	(3,3.18)

Out of these observations follows that Adam must have learned to form wh-questions at some point between 2,3 and 3,3. But when did he do this?

As we saw in table 1 as well as in Figures 4 and 5, there is no sudden improvement in Adam's performance, which points to the acquisition of a rule or a set of rules for the formation of wh-questions. Although the number of wh-questions increases towards the end of the study, Adam's improvement is small. Almost all of his questions over the entire period are grammatically correct, which means that he uses the appropriate wh-word, although there might be other errors such as a missing determiner or auxiliary. Errors such as, *What that right foot?, What that brought that?, What they did that?, etc.* are extremely rare in Adam's speech. Furthermore, uninverted wh-questions such as, *What I will write?, This is what?, Where my nest is?, etc.* are extremely rare as well. Most of the time Adam does not use an auxiliary in his questions and therefore hardly any inverted or uninverted questions can be found in the analysed data. And, as we have just observed, the majority of Adam's wh-questions throughout the entire period studied are formulaic.

Thus the question arises of how Adam's mental grammar at the age of 3,3 differs from the one at 2,3? In order to find out about the changes during this period, I will follow the development of specific formulas that Adam uses frequently. As we have seen earlier, the majority of Adam's questions are formed with *What, Where* and *Why*, which also show the greatest variety of formulas. I decided to follow questions consisting of a question word and two other words, which occur frequently so that they were categorized as major or

minor formulas. Furthermore, the formulas had to show a development. The most suitable formulas or formulaic frames were: *What* THING *doing?*, *Where* THING *going?* and *Why* THING PROCESS-*ing?*. Their development will be focused on in the following analysis. I will trace the formulas' development from formula to schema and categorize them into different stages of development.

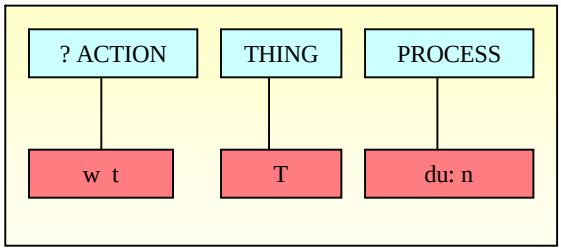
Example 1: What THING *doing?*

The formulaic frame *What* THING *doing?* appeared in Adam's transcripts at the age of 2,4.30 for the first time and was used the same way during the next two months. Thus this period can be called stage I and the formula is realized as given in the schema below.

Stage I (ca. 2,4-2,6)

What that paper clip doing?	(2,4.30)
What that train doing?	(2,4.30)
What that telephone doing?	(2,4.30)
What that egg doing?	(2,6.03)
What that cup doing?	(2,6.03)

What that THING ***doing?***



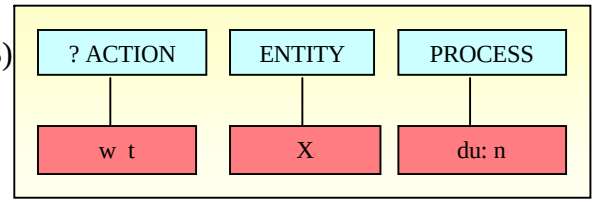
When looking at these examples, we can see how Adam forms his first questions. His first what-question was the formula *What that?*. Adam now uses this formula and extends it with a THING that performs an action. In Stage I the action is always expressed with the progressive *doing*.

In a second stage not only THINGs but also PERSONs perform the action of *doing* and also the *that* starts to disappear. Therefore the formula can be realized as follows:

Stage II (2,6-2,9)

What you doing (Mommy)?[15]	(2,6.03)
What are you doing?	(2,6.17)
What that button doing?	(2,6.17)
What the cowboy doing?	(2,9.04)

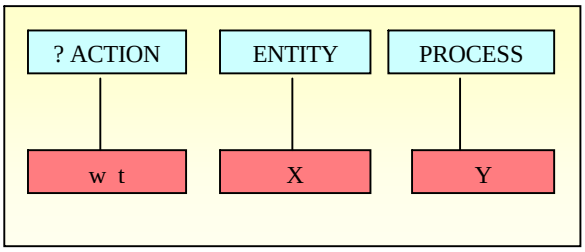
What ENTITY ***doing?***



After this variant of the formula was well established, Adam, in a third stage, begins to substitute other verbs for *doing* and the schema becomes more general. The phonetic form of /do: n/ now becomes the phonetic place holder Y.

Stage III (ca. 2,10-3,2) ***What*** ENTITY PROCESS-***ing?***

What he fixing? (2,10.16)
What he making? (2,10.16)
What Ursula doing? (2,10.30)
What you checking? (2,10.30)
What you looking for? (2,11.13)
What that boy doing? (3,0.11)
What the doggie doing? (3,1.09)
What that one doing? (3,1.26)
What you looking? (3,2.09)
What she pushing? (3,2.09)

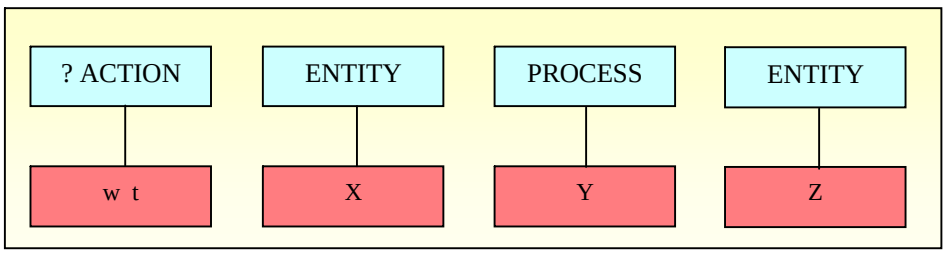


Finally, and considerably later than he has been able to vary the formulas and schemas of the previous stages, after he is able to exchange the subject and the verb, Adam manages to extend his formula and to abstract the schema further by adding another ENTITY. This development and examples are shown below in Stage IV.

Stage IV (ca. 3,3 +)

What she doing on her back? (3,3.18)
What you looking for? (3,3.18)
What she doing to her? [3] (3,3.18)
What the boy doing to her? (3,3.18)
What you doing looking at the furniture? (3,3.18)
What you doing with all those things in there? (3,3.18)

What **PERSON PROCESS-*ing* ENTITY**?



As we can see, there is s clear progress from the formula *What* THING *doing?* through increasingly more abstract formulaic frames (*What* ENTITY PROCESS-*ing*?) to a fairly general constructional schema in which none of the slots (except the question word) are tied to a specific lexical item (*What* ENTITY PROCESS ENTITY). Furthermore, the examples show that Adam's early usage of what-questions is highly stereotypical and gradually becomes more varied as new slots opened inside the formula.

Example 2: Where THING *going?*

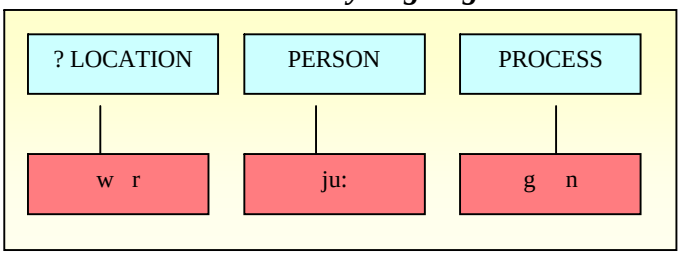
This formula was first recorded at the age of 2,10.2, considerably later and less frequently than the similar formula with *what*. This may be due to the limited recording times. We can be reasonably certain that Adam uses this formula more often when he is not recorded because, as pointed out earlier, formulas differ in frequency and need suitable situations and contexts. This is especially true for questions. We can only ask where someone or something is going when someone or something leaves the room, for example.

At the first stage, Adam uses the invariant formula *Where you going?* And only considerably later, at the age of 3,0.25 does he start to ask questions with subjects other than *you*. Again, the impression of this limited variety and late development may be due to the recording periods.

Stage I (ca.2,10-2,11)

Where you going? [5] (2,10.2)

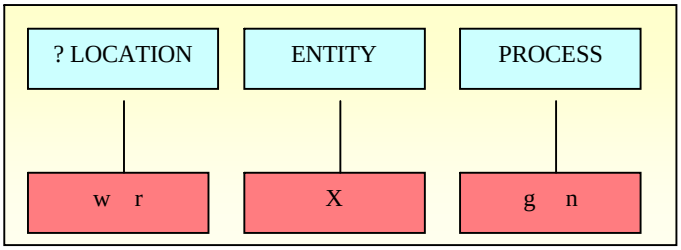
Where you going?



Stage II (ca.3,0 +)

Where* ENTITY *going?

Where she going? (3,0.25)
Where you going? (3,1.09)
Where they going? [3] (3,3.18)
Where the firemen going? [2] (3,3.18)



Example 3: Why THING PROCESS-*ing?*

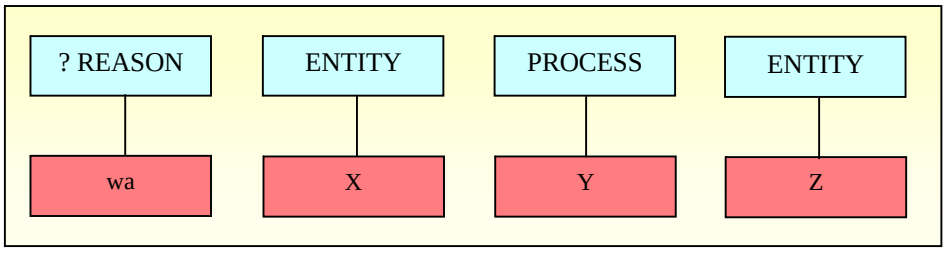
This formula appears in Adam's speech at the age of 2,10.30. There is no previous invariant formula or simple formulaic frame. The fact that Adam uses such a complex formula and fairly abstract constructional schema right away suggests that at some point in time he heard an utterance which was analysed as a formula but remained an unopened package. Only when Adam is 2,10, he opens this package and extracts a complex and abstract schema from it. Another possible explanation would be the limited time of recording. Adam has used this formula before but it just was not recorded. The answer to this phenomenon in Adam's why-question development is probably due to both of these facts.

Examples and the constructional schema are given below:

Stage I (ca. 2,10 +)

Why they fighting? [3]	(2,10.30)
Why you mixing baby chocolate? [2]	(2,10.30)
Why they making noise?	(2,10.30)
Why you waking me up?	(2,11.13)
Why he looking out window?	(2,11.28)
Why he driving the truck?	(3,0.25)

***Why* ENTITY PROCESS-*ing* ENTITY**?



4. Conclusion

The findings reported in this study, as well as the research done by Ewa Dobrowska, which was cited in the introductory section, strongly support the view that children's early questions are largely formulaic and the formula-to-schema hypothesis. It was argued that children reach adult-like productivity through the input they are given, by extracting schemas out of rote-learned formulas which they have analysed. There is no evidence for any mental mechanisms that install abstract combinatorial rules. As noted earlier, in Adam's as well as Naomi's speech, "there is no sudden improvement in performance which might signal rule acquisition."(98) Adam as well as Naomi rely on rote-learned formulas throughout the entire period studied.

Of course, Adam's command of English syntax improved very much during the one year studied. However, his progress should not be described as a movement towards more complex structures and accuracy but a movement towards more flexibility. Against a process towards more accuracy speaks that Adam's early usage of wh-questions is highly stereotypical and only gradually becomes more varied as new slots are open inside the formulas he uses. In addition, wh-questions with auxiliaries are mostly absent from his speech until the very end of the period studied. If there are auxiliaries in Adam's speech, they always appear in invariant formulas such as, *What is it?, Where is it? What are these?, etc.,* which are repeated over and over again. The reason for the absence of auxiliaries is, as shown earlier, that Adam has first acquired an invariant formula and only gradually learned to use it more flexibly. First he asks questions about different kinds of agents (*What* ENTITY *doing?*), then about locations (*Where* ENTITY *going?*) and finally about different kinds of actions (*What* ENTITY PROCESS-*ing* ENTITY?; *Why* ENTITY PROCESS-*ing* ENTITY?).

We have to realize that children's usage of language is very stereotypical and rather rigid, which suggests that they are relying on formulas and low-level schemas rather than abstract rules. Furthermore, children do not learn "abstract rules of the kind that linguists postulate to explain these structures [adult-like utterances], but a range of low-level schemas that have been gradually extracted from rote-learned formulas" (98).

Although this study confirms Dobrowska's findings, more children and data need to be studied in order to generalize the formula-to-schema hypothesis further. Furthermore, the entire corpus of Adam has to be studied in order to observe and analyse his further development and answer open questions such as the development of auxiliaries and their inversion in Adam's speech. This would also enable us to make an even closer comparison to Naomi's corpus, which was studied up to the age of 3,8. Already the studied period of one year reveals a very similar development of wh-question formation of Adam and Naomi, but at the same time there are differences. The stages Adam goes through in his development from formula to schema for the different question words do not completely match Naomi's stages. This, however is not problematic at all, since it was claimed that children acquire their competence and productivity out of the input they are given. This input differs from child to child. Therefore, it would be even more interesting to analyse input and the development of other children.

Appendix

Although, as just stated, the general trend in Adam's speech is towards more complex structures, the first formula he uses is often more complex than the following ones. For example, at the age of 2,3.4 he says: *What is in there, Mommy?* and only seconds later he reduces it to *What in there?* We also find constructions like *Where go?* (2,3.4)*; Where broom go, Mommy?* (2,3.18) before *Where?* (2,5.12) or *What that?; What happen, Mommy?* (2,4.3) before *What?* (2,4.15) All of Adam's wh-words are recorded in a construction before they appear on their own and sometimes this appearance is considerably later.

If Adam forms his questions by combining smaller units, we would expect the smaller units to appear before the more complex ones. However, this observed order of question development suggests that Adam has acquired a formula out of which he later extracts smaller units. Adam's development can be described as a u-shaped development. He first acquires as fairly complex formula, then he exacts smaller units out of it and comes back to the more

complex formula. As far as I can tell from the limited data, this u-shaped development seems to have occurred at different times in different formulas.[2]

References

Dobrowska, Ewa. "From Formulas to Schema: The Acquisition of English Questions." Cognitive Linguistics 11: 2000, 83-102.

Ingram David. First Language Acquisition: Method, Description, Explanation. Cambridge: Cambridge University Press. 1996. 436-444.

Ingram, David, Dorothy Tyack. "Inversion of Subject NP and Aux in Children's questions." Journal of Psycholinguistic Research 8: 1997. 333-341.

O'Grady, William. Syntactic Development. Cambridge: Cambridge University Press. 1997. 129-156.

[2] Cf. Ewa Dobrowska. 96/97.