Acquisition of exhaustivity in wh-questions: A semantic dimension of SLI?

Petra Schulz and Tom Roeper

This paper investigates how exhaustivity in single and multiple wh-questions is acquired in German-speaking children with SLI. Comparing semantic and pragmatic accounts of exhaustivity, obligatory exhaustivity of multiple wh-questions is argued to be problematic for pragmatic approaches. Thus, a unified semantic approach is suggested$chosen that relates exhaustivity to an inherent property of the question meaning. Two question-with-picture experiments, designed to favor exhaustivity, explored the comprehension of four wh-question types (single wh-questions with and without *alles*, paired and conjoined wh-questions) in 5-year-old children. Twenty children with SLI, 20 typically developing (TD) children, and 20 adults participated in Experiment 1, and 17 TD children in Experiment 2. The results indicate that 5-year-old TD children have aquired exhaustivity in single and paired wh-questions. The children with SLI mastered only the wh-*alles*-questions and performed at chance in the other wh-questions. For single wh-questions, the most frequent errors were singleton answers, and for paired wh-questions exhaustive lists of subjects or objects; plural responses were not found. Within individual children, single wh-questions were acquired before paired wh-questions. These findings suggest that a unified theory for both single and paired wh-questions is desirable attributing exhaustivity to universally exhausting the question domain, - a property that SLI children do not possess. These results add to recent research indicating that children with SLI may have deficits in semantics.

Keywords: SLI, wh-question, comprehension, German, exhaustivity, pair-list reading, single-list readings

1. Introduction

Wh-questions have been shown to be difficult for young typically developing (TD) children and to present persistent difficulties for children with Specific Language Impairment (SLI), both in production and comprehension. The majority of these studies focused on syntactic phenomena such as fronting of wh-words vs. in situ and the different structures of object and subject wh-questions, including the differences between *which*- and *who*-questions (cf. Friedmann and Novogrodsky, this volume, for an overview). To date, few acquisition studies have looked into the semantics and pragmatics of wh-questions. The present paper investigates how the exhaustivity property of single and multiple wh-questions is acquired by TD children and by children with SLI. More specifically, the experiments were designed to explore children's strategies in interpreting exhaustive wh-questions, rather than their knowledge of felicity conditions for posing such wh-questions.

In Section 2, arguing for the semantic approach, the semantic and pragmatic accounts of exhaustivity in single and multiple wh-questions are compared. Section 3 provides an overview of previous acquisition research on TD and SLI children's comprehension of exhaustive wh-questions across languages and states the Semantic Acquisition Hypothesis. The two comprehension experiments are described in Section 4 and Section 5. In Section 6 the results are discussed in light of the Semantic Acquisition Hypothesis, and it is concluded that the semantic account – providing a unified analysis of exhaustivity in single and paired wh-questions – can best capture the findings.

2. Exhaustive wh-questions in semantics and pragmatics

Wh-questions vary across languages in their syntactic and semantic properties (cf. Dayaal, 2005; Grohmann, $$$$; Hagstrom, 2003). For example, it is well-known that wh-questions may require fronting of all wh-words (e.g., Bulgarian), or fronting one wh-word (e.g., German, English), or may require the wh-words to be left in situ (e.g., Japanese). Moreover, in some languages multiple wh-questions are not allowed (e.g., Italian). Within the group of languages allowing multiple wh-questions, the possible readings differ. While in some languages, both exhaustive pair-list (PL) readings and single-pair (SP) readings are possible (e.g., Serbo-Croatian, Japanese), in others only PL interpretations are allowed (e.g., Bulgarian, German, English, Russian). Abstracting away from movement vs. no-movement accounts, generally it is assumed that for a wh-word to be interpretable it must involve two positions or syntactic objects, one serving as the operator and the other as the variable (e.g., for which x is it the case that x left?). Thus, wh-words – whether one or multiple – must create this operator-variable binding relation by the point of interpretation (cf. Hagstrom, 2003). One of the main questions is how to characterize this operator so that exhaustive readings can be derived in single and multiple wh-questions and how to account for non-exhaustive readings in single wh-questions.

This section sketches the (non)-exhaustivity in single wh-questions without overt lexical markers (Section 2.1.), and with lexical markers (Section 2.2.), in multiple wh-questions

(Section 2.3.), and in conjoined wh-questions (Section 2.4). The description of the pragmatic and semantic accounts of these facts (Section 2.5.) leads us to conclude that the semantic account is to be favoured (Section 2.6.). As cross-linguistic differences and parallels play an important role for the evaluation of semantic vs. pragmatic accounts, they are considered in some detail, even though the focus of the acquisition data is on German.

2.1. Exhaustivity in single wh-questions

Let us first consider the types of possible answers to a single wh-question REF _Ref232607823 \r \h (1) to REF _Ref234314361 \r \h (4) that have to be accounted for in a semantic or pragmatic account. As we are interested in the circumstances under which exhaustive answers are obligatory or not, unless stated otherwise the context is chosen such that more than one answer is possible.

A gave an exam to 20 students; Peter, Paul, and Mary failed. Her colleague B knows that not all students passed.

B: Who failed the exam?

A: Peter, Paul, and Mary.

A is at a workshop in Germany and is looking for directions to the city center. She addresses the non-local organizer B.

A: I need a person to tell me the way downtown. Who is a local?

B: The person in the white T-shirt.

A is in the lobby of a big hotel and asks the concierge B:

A: Where is a bathroom around here?

B: There's one down the hall, and one just one floor up.

A: Who is the pope?

B: A German.

Example REF _Ref232605281 \r \h (1) illustrates the mention-all reading, also referred to as 'exhaustive'. Both examples REF _Ref232607826 \r \h (2) and REF _Ref232610066 \r \h * MERGEFORMAT (3) demonstrate the mention-some reading, also referred to as non-exhaustive. More specifically, REF _Ref232607826 \r \h (2) demonstrates a 'singleton' answer and REF _Ref232610066 \r \h * MERGEFORMAT (3) a 'plural' answer. REF _Ref234314361 \r \h (4) presents a special case, as due to world knowledge the exhaustive answer is equivalent to the singleton answer. Thus, leaving aside the role of the context for these answer types and the question of a default reading, all three answer types – exhaustive, singleton, and plural –, are allowed in principle as responses to single wh-questions.

2.2. (Non)exhaustivity markers

Languages employ various expressions to explicitly mark a wh-question as non-exhaustive or exhaustive (cf. Zimmermann, 2007b, for an overview). First, expressions such as *for example* in English, *so* in German*, zoal* in Dutch (Beck and Rullmann, 1999) as

well as *su* in Hausa (Hartmann and Zimmermann, 2007) mark a non-exhaustive interpretation, more specifically a plural interpretation, as shown in the contrast between REF _Ref233021752 \r \h (6)a and REF _Ref233021752 \r \h (6)b as answers to REF _Ref233020579 \r \h (5).

a. Wer ist so zur Party gekommen? *German*

'Who for example came to the party?'

b. Su wàa suka zoo? *Hausa*

'Who for example came?'

a. Mary, Jane, and Sue. (out of 20 guests)

b. #? Mary. (out of 20 guests)

Non-exhaustivity markers thus block a singleton response and trigger a plural answer. Whether *so* and *for example* are also compatible with a exhaustive reading seems unclear (Zimmermann, 2007b).

Second, quantifying particles such as *alles* 'all' in German (Reis, 1992) *allemaal* in Dutch (Beck and Rullmann, 1999), *all* in Irish English (McCloskey, 2000), and *nèe* in Hausa (Hartmann and Zimmermann, 2007) function as exhaustivity markers. If added to a wh-question, the question is taken to be weakly exhaustive. REF _Ref233013808 \r \h (7)a and REF _Ref233013808 \r \h (7)b imply that all people who came to the party yesterday have to be identified and are thus incompatible with the singleton answer in REF _Ref233013808 \r \h (7)c. The position of *alles* is, as in Irish English, variable, with the

quantificational effects being the same (cf. REF _Ref233098366 \r \h (8)).

a. Wer alles kam gestern zur Party?

'Who all came to the party yesterday?'

b. Wàanee-nèe yá zoo? (Hausa)

'Who-EXH came?'

c. # Mary.

Wer (alles) kam (alles) gestern (alles) zur Party?

The semantic status of exhaustivity markers is still under debate. Reis (1992) and Hartmann and Zimmermann (2007) argued that particles like *alles* and *nèe* trigger a conventional implicature; Zimmermann (2007b) provided arguments for the presuppositional character of *alles*. Finally, it has been argued that *alles* possesses its own semantics, either operating directly on a question denotation and yielding a weakly exhaustive interpretation (Beck and Rullman, 1999, 288) or by operating on the meaning of the wh-expression by modifying the truth-conditions (Zimmermann, 2007a, in prep.). Following Zimmermann (2007a; in prep.), we assume as our working hypothesis that *alles* and *so* directly contribute to the semantics of the wh-expression. While *so* marks plurality, *alles* marks plurality and exhaustivity.

2.3. Exhaustivity in multiple wh-questions

While single wh-questions seem to be universally available, multiple wh-questions are not allowed in some languages (e.g., Italian, Irish, cf. Dayal, 2005). In languages that allow

multiple wh-questions, they can trigger either a pair-list (PL) or a single-pair (SP) answer (Dayal, 2005; Krifka, 2001). In English and German, multiple questions, also termed matching questions, presuppose that there is more than one answer and hence a SP answer is ruled out. For example, in a situation in which Jane, Tom, and Mary each ate something, the multiple wh-question REF _Ref233187406 \r \h (9)a requires an exhaustive PL answer as in REF _Ref233187406 \r \h (9)b. A SP answer is ruled out, as shown in REF _Ref233187406 \r \h (9)b'. Stress is marked by ', and infelicity of the answer is marked by #.

a. Whó ate whát?

b. Jane (ate) a banana, Tom (ate) a sandwich, and Mary (ate) a cookie.

b'. #Jane (ate) a banana.

Only in so-called REF-questions and echo-questions is the presupposition of the multiple wh-question absent that there is more than one answer, and hence a SP answer is allowed. REF-questions such as REF _Ref233187408 \r \h (10) trigger a SP answer.

a. A: Whó hit whóm first?

B: He hit me first.

b. A: Whó killed J.F.Kennedy whén? (cf. Krifka, 2001, 18)

B: Allegedly, Lee Harvey Oswald in 1963.

A PL answer is ruled out due to discourse knowledge – only two participants are present in REF _Ref233187408 \r \h (10)a – or due to world knowledge as in REF _Ref233187408

\r \h (10)b. Echo-questions as in REF _Ref233187412 \r \h (11) also call for a SP answer. Unlike REF-questions, echo-questions refer to a preceding utterance.

A: (almost inaudible) Jonathan ate green spinach.

B: Whó ate whát?

Thus, in contexts allowing for more than answer, in languages like English and German multiple wh-questions require an exhaustive PL answer.

Interestingly, to our knowledge, the answer type 'plural' we noted for single wh-questions has not been discussed with respect to multiple questions (but cf. Costa, 2004, for European Portuguese). Given that REF _Ref233535192 \r \h (9)b is a true description of the context, REF _Ref233191543 \r \h (12) is not a possible answer to the question *Who ate what?*

Jane (ate) a banana, and Mary (ate) a cookie.

Note, however, that a plural PL answer is available, if the marker *so* is added to one or both wh-expressions REF _Ref236236799 \r \h (13):

A: Mit wem (so) hat wer denn (so) gesprochen?

'With who (for example) did who (for example) talk?'

As shown in REF _Ref236324833 \r \h (13) *so* has the effect of cancelling the exhaustivity requirement in multiple wh-questions, similarly to its semantic contribution in single wh-questions, suggesting a uniform analysis of single and paired wh-questions.

Two further observations are pertinent to the discussion of exhaustivity of the PL answers in multiple wh-questions. First, the PL reading is blocked if one of the wh-words is inside a movement island (cf. Hagstrom, 2003). Second, the distribution of SP and PL readings is subject to cross-linguistic variation (Bošković, 2001; Hagstrom, 1998). In English, German, Bulgarian, and Russian multiple wh-questions like REF _Ref233535192 \r \h (9)a) are not felicitous in SP contexts, and hence do not allow SP answers like REF _Ref233187406 \r \h (9)b'). However, the SP reading is freely available in Serbo-Croation and Japanese (Grebenyova, 2006a,b). In other words, questions like *Who ate what?* are felicitous in both PL and SP contexts. A summary of the occurrence of multiple wh-questions across languages is given in REF _Ref233212600 \h * MERGEFORMAT Table 1. Interestingly, to our knowledge there are no languages of Type 3, allowing only SP readings but not PL readings for multiple wh- questions.

Grohmann (2003) reviews the Hagstrom-Boscovic approach to questions. According to that syntactic-semantic approach, wh-questions allow many variations in interpretation across languages that allow fronting of both of the wh-words. The semantic system involves a wh-particle which conveys a choice-function over wh-words that are in its scope. $$$

$Romanian (Grohmann, 2003)

Table SEQ Table * ARABIC 1

Availability of bare multiple wh-questions across languages

Context	Type 1 $ADD NAMES	Type 2	Type 3

	English, German, Russian, Bulgarian, Brazilian Portuguese, Romanian	Serbo-Croation, Japanese, Icelandic	?
Pair list (PL)	√	√	*
Single pair (SP)	*	√	√
Quiz context (Q-SP)	√	√ (by inference from the SP context)	√ (by inference from the SP context)

A caveat is in order regarding the findings summarized in REF _Ref233212600 \h * MERGEFORMAT Table 1. At this point it has to be left open whether the REF-readings in the quiz context, addressed for example by Krifka (2001) for German, may in fact be equivalent to the unmarked SP context, which is attested in Type 2 languages; that would result in merging Type 1 and Type 2 languages. Nevertheless, it seems reasonable to suggest that the exhaustive PL reading of multiple wh-questions, presupposing more than one answer, is the default reading across languages provided by UG (Grebenyova, 2006a,b; for German, cf. Krifka, 2001). In other words, Type 1 is taken to be the default option.

2.4. Conjoined multiple wh-questions

A different type of multiple wh-questions are conjoined wh-questions illustrated in REF _Ref233533169 \r \h (14). According to Krifka (2001), they do not presuppose a PL answer and hence are felicitous in a SP context.

A: Whó left and whén?

B: Máry left, at fóur.

The wh-words can also be conjoined and fronted. In German, this is possible for non-arguments only REF _Ref233533734 \r \h (15)a. As a response to REF _Ref233533734 \r \h (15)a a SP answer is preferred, while in REF _Ref233533734 \r \h (15)b a PL answer is favored (cf. also, Citko, 2008). In this regard, REF _Ref233533734 \r \h (15)a is parallel to the conjoined question REF _Ref233534818 \r \h (16).

a. Wann und wo hast du studiert?

when and where have you studied?

'When did you study, and where?'

b. Wánn hast du wó studiert?

when have you where studied?

'When did you study where?'

Wánn hast du studiert und wó?

when have you studied and where?

'When did you study, and where?'

Conjoined questions involve sluicing as shown in REF _Ref233620334 \r \h * MERGEFORMAT (17)a, where the ellipsis site is marked by strikethrough (e.g., Merchant, 2001). The structure underlying the second conjunct is assumed to be REF

_Ref233620334 \r \h (17)b, where governing an empty category in IP is allowed only for C^0 marked as a question via the feature [+Q] (cf. Kazenin, 2002).

a. When did you study and where ~~did you study~~?

b. When did you$_i$ study and [$_{CP}$ [$_{C[+Q]}$ where [$_{IP}$ ~~did pro$_i$ study~~]]]?

In Russian, according to Kazenin (2002) it is impossible to strand a wh-argument. This strict argument/adjunct asymmetry does not hold for German. In German stranded wh-expressions may not only be adjuncts REF _Ref233624484 \r \h (18)a but also arguments REF _Ref233624484 \r \h (18)b; both structures were tested in Experiment 1. In certain contexts the fronted wh-expression in the first conjunct may also be an adjunct REF _Ref233624486 \r \h (19).

a. Wer sitzt und wo?

who sits and where

'Who is sitting and where?'

b. Wer isst und was?

who eats and what

'Who is eating and what?'

Warum fährt Maria weg und mit wem/für wie lange?

why drives Maria awayand with whom/for how long

'Where is Maria going away and with whom/for how long?'

What seems crucial is that the first conjunct is grammatical by itself, i.e. all obligatory theta-roles have to be filled. Regarding their semantics, according to Krifka (2001, 22) conjoined questions are answered one question at a time, as shown in REF _Ref233685080 \r \h (20).

a. Who came, and when?

b. Question 1: Who came? Answer: Mary

c. Question 2: When did Mary come? Answer: At four.

d. Conjoined answer: Máry, at fóur.

Unlike in multiple wh-questions in Type 1 languages, a SP context is felicitous for a conjoined question such as REF _Ref233685080 \r \h (20). To our knowledge, PL contexts for conjoined questions have not been considered. Given the analysis in REF _Ref233685080 \r \h (20) for SP contexts, in a PL context (e.g., Mary came at four, John at five, and Sue at six), the conjoined question REF _Ref233685080 \r \h (20)a should not be answered with a conjoined answer like REF _Ref233685080 \r \h (20)d. We propose that two types of answers are felicitous: a double list answer as *Mary, John, Sue, - at four, five, and six,* and a PL response, resulting from repeating the procedure in REF _Ref233685080 \r \h (20)b-d, such as *Mary, at four; John, at five; and Sue, at six.*

$add FN: To the question "What happened" paired list answers are acceptable

2.5. Exhaustivity in wh-questions: Semantic and pragmatic accounts

The preceding sections showed that exhaustive and non-exhaustive readings are not equally available across various wh-question types. Distinguishing between the factors context (i.e., When is a wh-question is felicitous?) and answer (i.e., Which answers are felicitous in a given context?), we demonstrated that single, paired, and conjoined wh-questions as well as wh-questions with (non)exhaustivity markers are felicitous in contexts that allow for more than one answer. Moreover, it was shown that unlike single and conjoined wh-questions, wh-*alles*-questions, wh-*so*-questions, and paired wh-questions are infelicitous in contexts that only require one answer in Type 1 languages. REF _Ref236387799 \h * MERGEFORMAT Table 2 summarizes the types of answers that are available for the various types of wh-questions considered, given that the context allows for more than one answer.

Table SEQ Table * ARABIC 2

Felicity of answer types across different types of wh-questions (in contexts with more than one answer)

		Answer type	
Question type	Exhaustive	Singleton	Plural
Single wh	√	√	√
Wh-*alles*	√	*	*
Wh-*so*	(*)	*	√
Paired wh	√	*	*
Paired wh-*so*	(*)	*	√

Conjoined wh	√	*	*

Following Hamblin (1973), to know the meaning of a question is equivalent to knowing what counts as an answer. Then, how can these exhaustive and non-exhaustive answer types, depicted in REF _Ref236387799 \h * MERGEFORMAT Table 2, be derived? The proposals center around the questions of whether the distinction between mention-some and mention-all is a semantic or a pragmatic affair (e.g., Dayal, 2005; Groenendijk, 2008), and whether wh-questions are ambiguous or underspecified (e.g., van Rooy, 2004).

In the following we outline the two general proposals: the pragmatic account that we take to correspond to the underspecification view and the semantic account that we take to capture the view that wh-questions are ambiguous. In this paper we advocate the hypothesis that single and paired wh-questions should be captured in the same approach, the semantic one, and will suggest that the acquisition data are compatible with that approach.

According to the pragmatic account, the actual interpretation of a wh-question is underspecified by its conventional meaning (van Rooy, 2004; also Zimmermann, 2007a,b). Van Rooy (2004) proposes an operator *Op*(*P,w*), which rather than denoting the exhaustive or most informative value of a (n-ary) property *P* in a given world *w*, it denotes (one of the) optimal values of *P* in *w*. What constitutes an optimal value depends on the notion of relevance. The basic idea is to ask for the smallest set that gives optimal relevance. In this

line of reasoning, Zimmermann (2007a,b) assumes that wh-questions are semantically underspecified and the mention-all reading may come about as a result of a pragmatic interpretation strategy. According to Zimmermann (in prep.), *alles* then directly adds to the core propositional content of the question and its corresponding answers the exhaustivity that is not present in single wh-questions without *alles*. However, the fact that paired wh-questions and conjoined wh-questions (in more than one answer contexts) require exhaustive PL or double list answers, respectively, is in our view difficult to state in purely pragmatic terms. Assuming that the smallest set that gives optimal relevance is called for, without additional (possibly syntactic) restrictions it cannot be explained why multiple wh-questions have to be answered with an exhaustive PL.

According to the semantic account, wh-questions are ambiguous in that they receive different interpretations, arising from different underlying structures. Support for this view comes from the fact that many languages employ different wh-expressions to mark exhaustivity (cf. also Section 2.2). In English and German, for example, complex wh-pronouns like *which* plus number marking of the modified head noun express the requirement for a singleton or exhaustive interpretation (e.g., *Which students failed the exam? #Mary* vs. *Which student failed the exam? #Peter, Paul, and Mary*). If wh-questions are truly ambiguous, then also in the unmarked case of *who*-questions different structures should exist. This is exactly the approach proposed by Nelken and Shan (2004), which we adopted in the past (Roeper, Schulz, Pearson, and Reckling, 2007; Schulz and

Reckling, 2005). According to Nelken and Shan (2004), non-exhaustivity and exhaustivity of a wh-question are each reflected by a specific semantic representation. While non-exhaustive wh-questions involve existential quantification, as shown in REF _Ref232923043 \r \h * MERGEFORMAT (21), exhaustive wh-questions such as REF _Ref232923074 \r \h * MERGEFORMAT (22) involve universal quantification. To know who left in the (weak) exhaustive sense is to know, for each person x who left, that x left; this is illustrated in REF _Ref232923074 \r \h * MERGEFORMAT (22)c.

a. Where is a gas station?
   b. There is one just down the road.
   c. $\exists x.p$ (where p = it is common ground or known that p), $p = GSx$

a. Who left?
   b. Mary, Jane, and Sue left.
   c. $\forall x_n.\ p \rightarrow p,\ p = Lx$

Alternatively, within a structured meaning approach question meanings are understood as functions that when applied to the meaning of the answer yield a proposition (e.g., Krifka, 2001). As Krifka (2001) explicitly addresses the meaning of multiple and conjoined wh-questions, in the following we adopt the structured meaning approach. Wh-expressions such as *wer* 'who' denote atomic individuals and plural individuals, but not things or animals, as illustrated in REF _Ref236394781 \r \h (23):

[[who]] = {x | x ∈ *PERSON}

= {Mary, Jane, Sue, Mary+Jane, Jane+Sue, Mary+Jane+Sue, …}

Wh-questions denote structured propositions, where the wh-expression specifies the

focused question domain and the remainder of the question specifies the background of the structured proposition REF _Ref236394934 \r \h (24) (cf. Zimmermann, 2007a).

[[Who left?]] = < λx.x left, {x | x ∈ *PERSON}>

Informally speaking, the exhaustive reading is derived when the question domain is universally exhausted; a mention-some reading is derived from an existential quantification over the question domain. The exhaustivity feature could be modelled in different ways. Besides Nelken and Shan (2004), also Nishigaushi (1999) suggested that the wh-element in questions contains a variable that is quantificational in nature, - at least in languages like English. Additional support for the assumption that wh-pronouns host a hidden universal quantifier 'every' comes from a acquisition study by Heizmann (2008), who compared children's performance on universal quantifier structures such as *Is every farmer feeding a horse?* (requiring a no-answer, if an extra farmer is present) and exhaustive subject questions such as *What is under the table?* She found that children who fail to exhibit exhaustivity in wh-questions also fail to assign target-like scope to universally quantified statements. The notion that wh-questions are ambiguous is also compatible with the view that either the exhaustive or the non-exhaustive reading is the default (for the 'exhaustivity default' account, see Groenendijk and Stockhof, 1984; for the 'non-exhaustivity default', see Beck and Rullmann, 1999; Dayal, 1996, 2005; Hamblin, 1973; Kartunnen, 1977; Reis, 1992).

The presence of exhaustivity markers like *alles* and non-exhaustivity markers like *so* is

consequently explained as spelling out the different inherent meanings.

In the semantic account, the answer possibilities for multiple wh-questions, which differ according to context, are assumed to be derived from different semantic representations of these wh-questions. Note that the two facts sketched in Section 2.3, that the felicity of multiple wh-questions in SP contexts is subject to cross-linguistic variation, and that PL readings depend on grammatical restrictions such as islandhood, cannot be explained in pragmatic terms, given that contextual factors do not differ across languages in a systematic way. Extending Krifka's analysis (2001), we assume that the PL answers, which are obligatory in multiple wh-questions, are derived semantically as follows. Given that "… we can answer only one thing at a time" (Krifka, 2001, 21), multiple wh-questions are answered with only one 'thing' as well, namely via a function, i.e., a mapping procedure from a given and identifiable domain to values. Imagine a context in which Mary ate a banana and John ate an apple. The multiple wh-question REF _Ref233210432 \r \h (25) (that is asking for more than one thing at the time) is transformed by the operators in REF _Ref233210430 \r \h (26) to a question that asks for a mapping procedure (Krifka, 2001: 23):

Who ate what?$<\lambda<x,y>[\text{EAT}(y)(x)]$, $\text{PERSON} \times \text{THING}>$

a. $\text{FUN}(R) = \lambda f \forall x\ [x \in \text{DOM}(f) \rightarrow R(<x,f)x)>)]$, the set of functions f such that every x in the domain of of f stands in the R-relation to f(x)

b. $\text{FUN'}(A \times B)$ = the set of functions from A to B

The answer specifies a function by enumeration, such as *Mary banana and John apple* REF _Ref236410830 \r \h * MERGEFORMAT (27).

f: {Mary,John} → {banana,apple},

Mary → banana

John → apple

Conjoined wh-questions in PL contexts have to our knowledge not been formally characterized. Capturing the basic intuitions, questions such as *Who read, and what?* are assumed to be derived either parallel to the multiple wh-questions in REF _Ref233210432 \r \h (25) to REF _Ref236565742 \r \h (27) above, or as a double list answer, universally quantifying over each question domain, resulting in an answer such as *Mary and John, a banana and an apple*.

2.6. Summarizing exhaustivity in wh

Given a context with more than one possible answer, single wh-questions may be answered with a singleton, a plural, or an exhaustive list. Multiple wh-questions, however, in the unmarked case of a PL context require an exhaustive PL answer. Plural PL answers and SP answers are excluded. Only in a SP context such as a quiz situation is a SP answer felicitous. Addition of an exhaustivity marker like *alles* triggers exhaustive lists in single questions, and makes overt the exhaustivity of the paired lists in multiple questions. In contrast, the marker *so* triggers a non-exhaustive response, preferably a plural answer, in

both question types. Conjoined wh-questions are compatible with SP and PL contexts. In PL contexts, exhaustive answers are required. They can either be PL answers, as in the case of multiple wh-questions, or two separate lists, one for each wh-expression. Overall, exhaustivity occurs in both single and multiple wh-questions as do singleton/single pair answers, while plural answers are only attested in single wh-questions, corresponding to the mention-some reading. Under a pragmatic account, this absence of a plural PL answer in paired wh-questions has to be explained via stipulating different mechanisms for single and paired wh-questions.

Therefore, in the following, we assume a semantic account, according to which wh-expressions are ambiguous, with the exhaustive interpretation being the default in a more-than-one-answer context. The question to be asked then is whether the semantic account can account for the acquisition data.

3. Acquisition of exhaustivity in wh-questions

3.1. Comprehension of singleton wh-questions

Comprehension of singleton, i.e. non-exhaustive, wh-questions has been examined in several studies across languages, focusing on subject and object questions. In typical development, evidence is mixed with regard to the difficulty of these question-types. Some studies reported no difference between subjects and object-questions (e.g., Deevy and Leonard, 2004; Stromswold, 1995), while other studies found differences in performance

between subject and object questions (e.g., de Vincenzi, Arduino, Ciccarelli, and Job, 1999; Jakubowicz and Gutierrez, 2007). Similarly, in German, Siegmüller, Herzog and Herrmann (2005) reported a general advantage of interpreting subject over object questions and of argument over adjunct questions. Comprehension of argument wh-questions in singleton contexts was target-like at age 3 for TD children, whereas 2-year-olds performed at chance. Friedmann, Belletti, and Rizzi (2009) reported an advantage of *which*-subject questions over *which*-object questions for Hebrew$?, raising the possibility that the type of wh-pronoun play a role for interpretation.

Fewer studies focused on the comprehension of wh-questions in children with SLI (cf. Friedmann and Novogrodsky, this volume, for an overview). Exploring contrasts between subject and object wh-questions, and between *which*- and *who*-questions, studies mainly found deficits with *which*-questions and object questions in children with SLI (e.g., Ebbels and van der Lely, 2001; Friedmann and Novogrodsky, this volume). In a study with German-speaking children, the SLI group performed significantly better on subject than on object questions and, compared to their same age TD peer, performed significantly worse than the TD children in all question types (subject, object, adjunct) (Siegmüller et al., 2005; see also Penner and Kölliker Funk, 1998). Classifying children according to their performance on a standardized language test in two groups (low vs. at least average language abilities), Weissenborn, Höhle, and Penner (2006) found that at age 5 children in the low language ability group performed worse on single wh-questions than those the

average/above language ability group.

3.2. Comprehension of single exhaustive wh-questions

To date few studies have examined single exhaustive wh-questions in TD and SLI children. Experiments with typically developing children provided first evidence that the exhaustivity property of wh-questions is recognized by children between the ages of 3 and 6 (cf. de Villiers, 1995), with the age of mastery differing across languages, and possibly also according to argument-type. Following research by Roeper and de Villiers (1991, 1993) and Pèrez-Leroux (1993), Penner (1994, 1996) explored the bound-variable interpretation of subject wh-pronouns in Swiss-speaking TD children and found that the bound variable reading develops around age 4. Using subject questions in a controlled question-with-picture task (based on the question-after-story task from Roeper and de Villiers, 1991), Schulz and Penner (2002) found that 6- and 7-year old German-speaking TD children performed like the adult controls (age 6: 85%; age 7: 84%, adults: 98%). Their design, which was the origin of the two experiments presented in this paper, was as follows. Subjects saw a series of pictures, each depicting six individuals a subset of which exhibited the property in question, and heard a wh-question asking about the picture. This method we will refer to as $question-with-picture task. In the eight test trials, between two and five characters exhibited the property, while in the four control trials none or one of characters exhibited the property being asked about.

A cross-linguistic acquisition study of exhaustive wh-questions in English and German (Roeper, Schulz, Pearson, and Reckling, 2007; Schulz, Roeper, and Pearson, 2005) suggests that exhaustivity is acquired a year earlier in German than in English (age 5 vs. age 6). Notably, very few plural answers (6%) were found in either of the languages. This developmental advantage for German was also reported in a study with exhaustive object wh-questions (cf. Heizmann, 2007, 2008). Heizmann found adult-like performance in German-speaking children at age 4, and in English-speaking children at age 5. Hollebrandse (2003), using Schulz and Penner's (2002) design, found that the Dutch-speaking children between 4;4 and 5;6 performed worse than the German- and English-speaking children. Adults answered as expected, but only 26% of the Dutch children had mastered exhaustivity. Interestingly, while marking the verb for plural (*Wie <u>lezen</u> er een book?* Who read_PL expletive a book, instead of *Wie leest er een book?* Who read_SG expletive a book) significantly increased children's performance, still only 30% of the responses were exhaustive.

The two studies on exhaustive wh-question comprehension in children with language difficulties provided mixed evidence. In their 2002 study, Schulz and Penner also tested children with SLI and found that at age 6, they gave adult-like exhaustive answers in 41% of the cases and at age 7, in 62% of the cases. An analysis of the individual data verified that less than half of the children with SLI have mastered exhaustivity at age 6 (6 out of 16 children) and 7 (7 out of 16 children), compared to 81% of the TD children at both age 6

and 7 (13 out of 16 children each). The majority of errors consisted of singleton answers. Weissenborn et al. (2006) tested children at age 5;0 on single exhaustive wh-questions. Children in the low language group performed like average language-performers (cf. Section $$), with 64% correct responses. Only 10% of the errors were plural answers.

3.3. Comprehension of multiple wh-questions

Even though multiple wh-questions seem to be infrequent in the parental input (cf. Grebenyova, 2006a),$ADD: how often did they occu$$NF$ paired wh-questions have been noted to occur in children's speech around age 3. In a recent cross-linguistic study of the production of paired wh-questions in PL and SP contexts, Grebenyova (2006a, 2006b) reported that 4-year-old English- and Russian-speaking children, like the adults, produced multiple wh-questions in PL, but not in SP contexts. Children acquiring Malayalam, a language that allows multiple wh-questions in SP contexts, in contrast also produced these questions in SP contexts, even though to a lesser degree than the adults tested (14 vs. 44%, cf. Grebenyova, 2006a, 183).

Only few studies have looked into the comprehension of multiple wh-questions, all of which explored paired wh-questions of the type *Who is eating what?* and none of them conjoined wh-questions of the type *Who is eating and what?*. Using a question-with-picture task, Roeper and de Villiers (1991) found that while adults consistently responded with paired exhaustive lists, four- to six-year-olds responded with exhaustive paired lists in

78% of the cases, and younger children showed mastery of exhaustive paired lists in only 32% of the cases. This developmental pattern was confirmed for German by Heizmann (2008). The number of PL responses increased with age, with about 20% PL answers at age 3 and 90% PL answers at age 5. Interestingly, in English using the same task mastery of multiple wh-questions occurred later, with only 60% PL responses at age 7. Both German- and English-speaking children's incorrect responses were mostly exhaustive object lists. Based on data from a standardized test (DSLT, Seymour, Roeper, and de Villiers, 2000), Roeper (2004) reported that children with language impairment have persistent difficulty understanding multiple wh-questions and performed significantly below their same-age peers until age 9.

3.4. Summary of the acquisition studies

Taken together, mirroring the situation in theoretical semantics/pragmatics, apart from Heizmann (2008) for TD acquisition, comprehension of single and multiple exhaustive wh-questions has so far been studied independently. Based on her finding that single wh-questions are acquired before multiple wh-questions, Heizmann (2008) proposed that this asymmetry results from the number of sets that have to be established. In single wh-questions, only one set has to be checked for relevant properties of its members. In multiple wh-questions, two sets have to be established and related to each other, e.g., in *Who is eating what?* the set of subjects and the set of objects being eaten. However, as only group

data are given, it remains open whether the proposed developmental path is also found in individual children. The few studies on comprehension of single and multiple exhaustive wh-questions in SLI suggest that children with SLI perform significantly below same-aged typically developing children. As single and paired wh-questions were not tested within the same study, the relation between the acquisition of both question types remains open for SLI as well, especially because of the heterogeneity of deficits observed across individual children with SLI. To our knowledge, the interpretation of overt exhaustivity markers in wh-questions and of paired wh-*und*-questions has not been studied in acquisition.

We regard the difficulties of SLI children with exhaustivity to be a strong indicator for a general problem with quantification that SLI children face and that may in fact be an underappreciated special kind/subtype of disorder (cf. Roeper, 2007; Schulz, 2010).

3.5. Research questions and hypotheses

The main goal of this study is to compare the comprehension of exhaustive single and multiple wh-questions in SLI and typical development and to account for the intermediate steps in the acquisition paths. To determine the exact locus of the expected difficulties with these structures in children with SLI, two types of wh-questions were included that have to our knowledge not been tested before in acquisition and that present a less complex variant of the single and multiple wh-question: wh-questions with the overt exhaustivity marker *alles*, where an exhaustive interpretation does not rely on a covert operator, and conjoined

wh-*und*-questions, which require exhaustivity in the answer list, but where PL answers are not obligatory.

Which acquisition hypotheses can be derived from the semantic account of wh-questions? As exhaustivity in single and multiple wh-questions has mostly been considered independently, the following considerations are necessarily speculative. Under the semantic view, exhaustivity is a feature represented in the structure of both single and multiple wh-questions. Overt markers such as *for example* and specific contexts (e.g., quiz context in the case of multiple wh-questions) relax the exhaustivity requirement. Extending Krifka's (2001) account, exhaustivity is formalized as a universal quantifier exhausting the function domain, which varies depending on the type of wh-pronoun, e.g., PERSON for *who* or THING for *what*. Consequently, in multiple wh-questions, a function f over pairs is required that exhausts the domain of f, resulting in a more complex structure than in single exhaustive wh-questions. Exhaustive answers rest on the discovery of the relation between universal quantifier and the function domain, whereas plural answers hinge on encountering contexts in which an *for example*-answer is called for. From this line of reasoning, the Exhaustivity Acquisition Hypothesis in REF _Ref236308529 \r \h (28) can be formulated:

Exhaustivity Acquisition Hypothesis

a. Exhaustivity in multiple wh-questions is recognized later than in single wh-questions ($for complexity reasons)

b. Before mastery of exhaustivity, wh-pronouns are interpreted as a constant, resulting in singleton responses.

c. Plural responses are not an intermediate acquisition step; they are acquired only after exhaustivity is mastered.

$alaborate %FM: Ad REF _Ref236308529 \w \h (28)b the difference between constant and existential: if exitanetial, mention-some answers should be ok, i.e. also plurals..

Under a semantic account, wh-*alles*-questions should be acquired early, as *alles* carries its own exhaustivity feature. Conjoined wh-questions, forcing exhaustivity but not PL answers, should be acquired at the same age as single exhaustive wh-questions.

Regarding the core of the deficit in SLI, we assume that children with SLI are delayed in the acquisition of exhaustivity, but in principle follow the same acquisition path as TD children (cf. de Villiers, 2003; Leonard, 1998: for discussion of delay vs. deviation). Consequently, the Exhaustivity Acquisition Hypothesis pertains to both TD and SLI. The specific difficulty in children with SLI is stated in REF _Ref236216898 \r \h (29) (modifying previous versions, cf. Roeper, 2009; Schulz and Reckling, 2005).

Missing quantifier hypothesis

Children with SLI do not recognize that the question domain has to be exhausted.

$1: Note that this difficulty predicts a delay in acquisition, and not a deficit, as TD children are assumed to start out without providing exhaustive answers as well. However, we leave

open at this point whether children with SLI will arrive at the same target grammar as TD children, supporting the delay assumption, or whether $$. (but cf. Schulz, 2010, for a deficit account).

$2: A reviewer raises the interesting question whether we refer to a semantic or syntactic deficit or all types of SLI. As research on difficulties with semantic/pragmatic phenomena is just beginning to emerge, we refrain from a more refined characteristic of the type of SLI involved. It may be that many children with syntactic difficulties also show problems in exhaustivity or that only children with semantic difficulties show an exhaustivity problem (Schulz, 2010).

4. Experiment 1

$As a starting Point, only who. .no D-linked wh-expressions. $ This experiment explored the comprehension of exhaustive wh-questions in children with SLI and with typical development. To detect non-adult patterns, we chose a pragmatic context that would evoke consistent exhaustive responses in adults. This way, if children with SLI or typical development fail to supply exhaustive answers, this could not be attributed to a context favoring non-exhaustive answers in general.

4.1. Participants

We tested 20 5-year-old typically developing children (mean age = 5;8; range = 5;0 to 5;11; SD = 3,5 months; 14 boys) and 20 5-year-old children with SLI (mean age = 5;4, range = 5;0 to 5;10, SD = 3,1 months; 10 boys). In addition, 20 adults served as a control

group (age range = 25 to 57). They were tested individually, following the same procedure and using the same material as the children, except for the SETK 3-5 and 2 pretests.

The children in the TD group all attended regular kindergartens and were reported to not show any signs of language, speech, or hearing impairment. The children with SLI were enrolled in special language impairment intervention programs at their kindergartens. All the children in the SLI group met the following exclusionary criteria for SLI (Leonard, 1998): They had no hearing impairment and no recent episodes of Otitis Media; they showed no evidence of obvious neurological impairment or impaired neurological development; and they had no symptoms of impaired social interaction that are typical of autism. All the children in the SLI group had been diagnosed with SLI prior to the study through clinical assessment by a speech-language therapist, based on non-standardized tests used in the clinics.

4. 1.1 Performance in the standardized language test SETK 3-5

Inclusion in the TD or SLI group was based on children's performance in the standardized language test SETK 3-5 (Grimm, 2001), which contains five subtests aimed at diagnosing children with SLI. Two subtests assess morpho-syntactic abilities: Using an act-out-task, Subtest VS assesses comprehension of sentences in varying complexity, and in an elicited production task, Subtest MR tests knowledge of plural marking. Three subtests assess memory related abilities, using elicited imitation: Working memory is assessed via non-word repetition in Subtest PGN, via real words in GW, and with sentences in increasing complexity in subtest SG. The inclusion criterion for the TD group

was performance at average or above in at least 4 out of 5 subtests of the SETK 3-5, and for the SLI group below average performance in at least 2 out of 5 subtests. All subjects met the inclusion criteria for participating in the experiment (cf. Appendix A. for the detailed test results). Sixteen out of 20 TD children performed well on all 5 subtests of the SETK 3-5; and 4 TD children had T-values < 40 in just one subtest (VS, MR, PGN). All 20 children with SLI failed in at least 2 subtests, eight failed in 3 subtests, and three SLI children failed in 4 subtests. Out of the 20 children with SLI, 15 performed below average in VS, 5 in MR, 18 in PGN, 1 in GW, and 15 in SG, pointing to great difficulties in sentence comprehension and with working memory for non-words, but not with plural morphology. SLI children's performance was significantly poorer than that of the TD children in all subtests of the SETK 3-5, VS: $t(38) = 6.36$, $p < .001$; MR: $t(38) = 4.56$, $p < .001$; PGN: $t(38) = 8.15$, $p < .001$; GW: $t(38) = 4.22$, $p < .001$; SG: $t(38) = 10.53$, $p < .001$.

4.2. Design

4.2.1 Question-with-picture task: $materials, procedure, coding

This experiment employed the question-with-picture task. The experimenter showed the child a picture introduced by a short lead-in sentence, and then asked a wh-question, while the child was looking at the picture. Each of the participants was tested individually in a quiet room in two sessions; the sessions were ca. 4 weeks apart. In session 1, children received two subtests of the SETK 3-5 (VS, SG), a wh-pretest and a vocabulary pretest,

and the first part of the main experiment. In session 2, children were administered the remaining three subtests of the SETK3-5 (PGN, MR, GW) and the second part of the main experiment. All sessions were video-recorded for later data check against the onsite-coding and for further individual analyses. No response-contingent feedback was given by the experimenter. When the child failed to supply an answer, items were repeated once.

At total of 26 wh-questions, 20 test items and 6 control items, were presented to each child. There were 4 conditions, each comprising 5 test items, which were presented to each participant. Condition one contained a single wh-pronoun (henceforth single wh-question), condition two contained a single wh-pronoun and the lexical exhaustivity marker *alles* (henceforth single wh-*alles*-question), condition three contained two wh-pronouns (henceforth paired wh-question), and condition four contained conjoined wh-questions with two wh-pronouns (henceforth paired wh-*und*-question). All single wh-questions used the same wh-pronoun *wer* 'who' to achieve comparability across items. The paired wh-questions were construed with a fronted subject wh-pronoun *wer* 'who' and an object or adjunct wh-pronoun (*wen* 'whom', *was* 'what', *wo* 'where', *mit was* 'with what') in situ. In order for the wh-question to be felicitous, a verbal discourse context was created for each item. It was kept minimal, however, to ensure that the child was not prompted by explicit remarks about the individuals shown on the picture to give a non-singleton answer. Typical test items for single wh-questions and single wh-*alles*-question are illustrated in REF _Ref232694265 \r \h (30) and Fig. 1.

Guck mal, was ist denn hier los? 'Look, what is happening here?'

a. Wer hat einen Fußball?
   who has a soccer ball
   'Who is holding a soccer ball?'

b. Wer hat alles einen Fußball?
   who has all a soccer ball
   'Who (all) is holding a soccer ball?'

Insert REF _Ref231205684 \h Figure 1 about here

For paired wh-questions, a sentence describing the scene was added, without referring to single individuals or the activities. Typical test items for paired wh-questions ad paired wh-*und*-questions are given in REF _Ref236337499 \r \h (31) and Fig. 2.

Die haben Hunger. 'They are hungry.'

a. Wer isst was?
   who eats what
   'Who is eating what?'

b. Wer isst und was?
   who eats and what
   'Who is eating and what (is he eating)?'

--
Insert REF _Ref231205701 \h Figure 2 about here
--

Each picture displayed several family members, which were introduced in an initial picture as a family, naming them as *mother, father, grandfather, grandmother, boy, girl* to minimize memory effects. In the single wh- and the single wh-*alles*-condition, there were always six individuals out of which between two and five shared the property being asked about, such as holding a soccer ball. This variation ensured that children could not develop guessing strategies, such as listing all individuals on the picture or consistently responding with the same number of individuals. The four control items required a singleton answer and served to prevent the child from assuming that she always had to respond with more than one individual. In the paired wh- and paired wh-*und*-condition, the pictures displayed between two and four individuals sharing the same property such as eating something. Note that in order to reduce the processing load required for giving paired list answers, the maximum number of individuals was four instead of six. The two control items depicted

two individuals, one of which was engaged in the activity being asked about, and required a SP answer. Even though multiple wh-questions presuppose a list answer, these items were included to prevent the child from assuming that she always had to respond with more than one pair. The verbs used in the stimuli met the following criteria: They were easy to illustrate within a one-picture set up, and they were part of the lexicon of preschoolers (e.g., *drink, eat, read*, *sit*).

The items in the four conditions were presented in a block design to minimize carry over effects from presence of the lexical exhaustivity marker *alles* and the conjunction *und* to the other conditions. Within a block, test and control items were presented in a random but fixed order. Four different test versions were created, in which order of the conditions was varied but beginning with single wh-questions was kept constant.

4.2.1 Pretests

The main experiment was preceded by two pretests. In Pretest 1, children's comprehension of single non-exhaustive wh-questions was assessed, using the question-with picture task with four items from the Penner Screening (1999). A picture was described with a short sentence (e.g., *Reto und sein Vater machen einen Kuchen* 'Reto and his father are making a cake'), followed by a wh-question, (e.g., *Mit wem macht Reto einen Kuchen?* 'With whom is Reto making a cake?)'. This way it was ensured that both the TD and the SLI group are able to interpret single non-exhaustive wh-questions, with a target-like performance defined as 80% correct or above. Pretest 2 assessed children's knowledge of the vocabulary used in the main experiment to ensure that interpretation of the wh-questions was not precluded by lexical deficits. The vocabulary test included 34 pictures of the objects used in the main experiment (e.g., soccer ball, chocolate). The cards were placed on the table upside down, and the experimenter asked the child to turn over one card at a time and tell her what she saw. If the child did not know the word, the experimenter supplied the correct word and placed the card on the table again. Failure to name an object a second time was coded as incorrect. Passing criterion was set at 30 out of 34.

Participants' responses were coded as correct or incorrect. REF _Ref231185698 \h * MERGEFORMAT Table 3 summarizes the different types of correct responses (for a detailed analysis of the error types, cf. REF _Ref231185111 \h * MERGEFORMAT

Paired wh-und (*Who is eating what and?*)

Table 4 and REF _Ref231185129 \h * MERGEFORMAT Table 5). Note that the coding for the responses to wh- and wh-*alles*-questions was the same.

Table SEQ Table * ARABIC 3

Types of correct responses to the test items

Condition	Response type	Example
Wh(-alles) (*Who has a soccer ball?*)	Verbally listing the subjects with VPs	*The dad has a soccer ball, the child has a soccer ball, and the boy has a soccer ball*
	Verbally listing the subjects	*The dad, the child, and the boy*
	Pointing to the subjects	*This, this, and this one* (points)
	Combination of pointing and verbal response	*This, this, and this one* (points), *this girl, the boys, and the dad*
	Listing by exclusion	*Everybody but these two*
Paired wh (*Who is eating what?*)	Verbally listing pairs with verb	*The sister eats a banana, the boy chocolate, the grandma an apple, the grandpa a fish*
	Verbally listing pairs	*The boy chocolate, the girl banana, the grandma apple, the grandpa fish*
	Pointing and verbally listing the pairs	*The boy chocolate, the girl banana, the grandma apple, the grandpa fish* (points to the subjects while speaking)

Paired wh-und (*Who is reading and what?*)	Listing subjects and objects separately	*Grandpa and Mom, the newspaper and a book*
	Grouping subjects, listing objects separately	*All, the newspaper and a book*
	Verbally listing pairs with verb	*Mom reads the newspaper, and Grandpa reads a book*
	Verbally listing pairs	*Mom the newspaper, and Grandpa a book*
	Pointing and verbally listing the pairs	*Mom the newspaper, and Grandpa a book* (while pointing to the subjects)

Importantly, correct answers to paired wh-*und*-questions are of two different types: responding with two separate lists for subjects and objects or responding with a list of pairs. We decided for this coding option for two reasons. First, a detailed account of the semantics of paired wh-*und*-questions that would justify the rejection of paired responses on theoretical grounds is still missing. Second, PL answers are allowed in this condition, as also confirmed by the adults' preference for PL answers in the current experiment (93%).

4.3. Design, materials, procedure

4.3. Results

4.3.1$. Pretests

The comprehension of single non-exhaustive wh-questions (Pretest 1) was good for both children with SLI and TD children, with an average of 85% correct (SD = 20.52) for the SLI group and 95% correct (SD = 13.08) for the TD group. A non-parametric Mann-

Whitney test yielded an almost significant difference between the two groups, $U = 148.0$; $p = .073$. Given the fact that children with SLI often suffer from language deficits in multiple areas of grammar and language comprehension, it is not surprising that the TD children outperformed children with SLI in this task. As this pretest aimed at establishing that the SLI group did not suffer from a general deficit in the comprehension of non-exhaustive wh-questions, all children participated in the main experiment.

The results of Pretest 2 were similar to the results of Pretest 1. Both children groups performed at ceiling in the active vocabulary test, with an average of 97.2% correct (SD = 38.93) for the SLI group and 100% correct for the TD group. Analysing the individual data, all children met the criterion of naming at least 30 out of 34 items. Nine SLI children named 34 items, seven SLI children named 33 items, one SLI child each named 32 and 31 items, respectively, and two SLI children named 30 items correctly.

4.3.2. Main experiment

Based on the classification of responses as correct in REF _Ref231185698 \h Table 3 adults' performance was at ceiling (100% correct) in all test conditions and in the 2 control conditions. Therefore, their data were not considered any further in the quantitative analysis.

An analysis of variance was (ANOVA) was performed for the child participants over the percentage of correct answers with type of wh-question as within-subjects factor and with test version and group as between-subjects factors. There was a main effect of group, F (1, 32) = 28.46, $p < .001$, but not of test version, F (3, 32) = 2.15, $p = .114$. The

interaction of group by test version was not significant, F (3, 32) = 1.73, p = .181. Therefore, in the following data were collapsed across the different test versions. A second ANOVA was performed for the child participants over the percentage of correct answers with type of wh-question as within-subjects factor and group as between-subjects factor. There was a main effect of group, F (1, 38) = 24.79, p < .001, and of type of wh-question, F (3, 38) = 11.04, p < .001. The interaction of group by type of wh-question was significant, F (3, 38) = 4.89, p = .003.

The main finding of this experiment was that, compared to the five-year-old TD peers, the children with SLI had considerable difficulty understanding exhaustive single and paired wh-questions. The results for the TD and the SLI group are summarized in Figure 3.

Insert REF _Ref231205737 \h * MERGEFORMAT Figure 3: about here

---

The results show that only performance on single wh-*alles*-questions was as high as in the TD group (SLI: M = 91, SD = 22.3 vs. TD: M = 100, SD = 0). SLI children showed poor performance on single wh-questions (M = 75, SD = 39.9), and even lower performance both on paired wh-questions (M = 49, SD = 37.1) and on paired wh-*und*-questions (M = 48, SD = 41.7). T-tests were used for all paired comparisons between the TD and the SLI group, except for wh-*alles*-questions, where TD children's performance was at ceiling. Children with SLI performed significantly poorer than the TD group in the remaining 3

conditions: comprehension of single wh-questions, $t(38) = 2.24$, $p = .031$, comprehension of paired wh-questions, $t(38) = 4.70$, $p < .001$, and comprehension of paired wh-*und*-questions, $t(38) = 4.06$, $p < .001$.

Looking at the TD group, their performance on single wh-questions was significantly better than their performance on paired wh-questions, $t(19) = 2.35$, $p = .030$. The other possible comparisons did not yield significant differences.

Regarding the SLI group, the difference between single wh-questions and single wh-*alles*-questions was not significant, $t(19) = 1.80$, $p = .088$, and the difference between single wh-questions and paired wh-questions approached significance, $t(19) = 2.00$, $p = .060$. Performance on single wh-questions and paired wh-*und*-questions differed significantly, $t(19) = 2.38$, $p = .028$, while performance on paired wh-questions and paired wh-*und*-questions did not differ significantly in the SLI group, $t(19) = .13$, $p = .895$.

Comprehension of the control items was mostly target-like for the TD and the SLI group. Performance on single wh- and single wh-*alles* controls, requiring a singleton answer, was at ceiling for both the children with SLI and with TD. TD children performed significantly better on paired wh-controls than SLI children (TD: M = 92.5, SD = 17.9; SLI: M = 57.5, SD = 42.6), $t(38) = 3.3$, $p < .01$. This difference is expected as the controls in this condition require a SP response.

Error analysis. Children's incorrect responses were then grouped according to error type to more closely examine the source of children's errors. For single wh-questions, the

following error types were observed: naming more than one but not all required individuals (Plural), and naming one individual only (Singleton). As can be seen in REF _Ref231185111 \h Table 4 below, most of SLI children's errors in single wh-questions are singleton responses.

Table SEQ Table * ARABIC 4

Experiment 1. Types and number of errors for single wh-questions by subject group (number of the participants who committed this error in parentheses)

Group	Condition	Errors total	Error types (Number of participants)
			Plural Singleton
TD	wh	3 of 100	3 (3)	-
	wh-*alles*	-	-	-
SLI	wh wh-*alles*	25 of 100 9 of 100	2 (2) 2 (2)	23 (5) 7 (3)

For paired wh-questions the following types of incorrect answers were found: naming more than one but not all required pairs (Plural pairs), naming one pair (1 Pair), responding with an exhaustive list of subjects (Subj-list) or of objects (Obj-list), naming one subject (1 Subj) or one object (1 Obj), and a small number of further types of incorrect answers

(Other). The distribution of error types to paired wh-questions is given in REF _Ref231185129 \h * MERGEFORMAT Table 5 below.

Table SEQ Table * ARABIC 5

Experiment 1. Types and number of errors for paired wh-questions by subject group

Group	Condition	Errors Total	Error types
Plural pairsSingle PairSubj-listObj-list1 Subj1 ObjOtherTDpaired wh9/100--33--3*paired wh-*und*8/1001-4-2-1Total17/2001-732-4SLIpaired wh51/100-82312-26paired wh-*und*52/100-819+9763Total103/200-164221789* saying *alle/jeder;* + some of them including the verb			

Notably, plural pair responses basically do not occur in either group. Responding with an exhaustive list of subjects is attested in the TD group, and accounts for almost half of the errors in the SLI group. Other recurrent errors among the children with SLI are answers that consist of an exhaustive list of objects or of a single pair.

Analysis of individual responses. Individual responses were calculated to investigate whether the observed group differences between TD and SLI children were also found in children's individual performance. REF _Ref231205837 \h Table 6 shows the percentage of correct responses to the four test conditions for each child in the two subject groups.

Table SEQ Table * ARABIC 6

Experiment 1. Number of subjects by responses correct for each condition according to subject group

Number of correct responses	Single wh		Single wh-*alles*		Paired wh		Paired wh-*und*	
(5 correct responses possible per subject)	TD	SLI	TD	SLI	TD	SLI	TD	SLI
0/5	0	4	0	1	0	4	1	7

1/5	0	0	0	0	0	5	0	1
2/5	0	1	0	0	0	0	0	2
3/5	0	0	0	0	1	4	0	3
4/5	3	1	0	4	7	3	3	1
5/5	17	14	20	15	12	4	16	6
Mastery (4 or 5 responses correct)	20	15	20	19	19	7	19	7

Chi-square analyses comparing the individual responses among the TD and the SLI group for each of the four conditions showed that the distribution of responses differed significantly for comprehension of paired wh-questions, $\chi^2(5, N = 40) = 16.400, p < .01$, and comprehension of paired wh-*und*-questions, $\chi^2(5, N = 40) = 16.045, p < .01$, but not for comprehension of single wh-questions and single wh-*alles*-questions.

Mastery of a wh-question-type was defined in a strict way as providing at least 4 out of 5 correct responses in order to assess consistent use of the target-like response type. According to this definition, almost all of the five-year-old TD children had mastered exhaustive wh-questions in all four conditions (cf. REF _Ref231205837 \h Table 6). In contrast, among the children with SLI, 19 out of 20 had mastered single wh-*alles*-questions, 15 out of 20 (75%) responded as though they had mastered single wh-questions, and seven out of 20 (35%) had mastered paired wh-questions and paired wh-*und*-questions, respectively.

Regarding the paired wh-questions and the paired wh-*und*-questions, a variety of

individual error patterns was observed. Most notably, the single pair-error, being observed 8 times each across children, was not the favoured strategy of any of the children. Out of the 16 children who did not answer all paired wh-questions correctly, six provided correct paired answers along with exhaustive subject and/or object lists. Out of the 14 children who did not answer all paired wh-*und*-questions correctly, five provided correct paired answers together with exhaustive subject and/or object lists.

In a next step, the individual response patterns of the SLI group across conditions were examined to discover possible dependencies between mastery of single and mastery of paired wh- and paired wh-*und*-questions, respectively. REF _Ref231205872 \h Table 7 and REF _Ref231205922 \h Table 8 illustrate the number of children who mastered single wh-question and one of the other question types. Note that w*h-alles*-questions were excluded from the analysis as SLI children's performance was near ceiling.

Table SEQ Table * ARABIC 7
Experiment 1. Number of children with SLI by mastery of simple wh-questions and paired wh-questions

			Paired wh-questions	
		No mastery	Mastery	Total
Single wh-questions	No mastery	3	2	5
	Mastery	10	5	15
	Total	13	7	20

Table SEQ Table * ARABIC 8
Experiment 1. Number of children with SLI by mastery of simple wh-questions and paired wh-*und*-questions

			Paired wh-*und*-questions	
		No mastery	Mastery	Total

Single wh-questions	No mastery	4	1	5
	Mastery	9	6	15
	Total	13	7	20

As can be seen in REF _Ref231205872 \h Table 7 above, half of the SLI children had mastered single exhaustive wh-questions, but not paired wh-questions. Out of 7 SLI children who showed mastery of paired wh-questions, only two had not mastered single wh-questions. REF _Ref231205922 \h Table 8 above confirms this picture. Nine out of 20 SLI children (45%) had mastered single exhaustive wh-questions, but not paired wh-*und*-questions, and out of seven SLI children who had mastered paired wh-*und*-questions, only one did not show mastery of single exhaustive wh-questions.

Apart from mastery, the test conditions allow for inspecting the response type 'exhaustive list' across conditions. SLI children who failed paired wh-questions by systematically responding with exhaustive lists of either subjects or objects, should make use of this pattern in their responses to the single wh-questions as well. REF _Ref231207248 \h Table 9 and REF _Ref232694899 \h Table 10 illustrate the number of children who exhibit exhaustivity across question types. The response pattern 'exhaustivity' is attested if at least 4 out of 5 responses fall into this category, i.e. exclusively name all x.

$alternative: make the tables only for those who failed paired questions??

Table SEQ Table * ARABIC 9
Experiment 1. Number of children with SLI using the exhaustive list strategy in simple wh-questions and paired wh-questions

			Paired wh-questions	
		No exhaustive lists	Exhaustive lists	Total
Single wh-questions	No exhaustive lists	4	**1**	5
	Exhaustive lists	12	3	15
	Total	16	4	20

Note that the majority of children (16/20) did not use exhaustivity as a systematic response strategy in paired questions. Out of the four children who did, only one did not use this response type with single wh-questions (JA).

Table SEQ Table * ARABIC 10

Experiment 1. Number of children with SLI using the exhaustive list strategy in simple wh-questions and paired wh –*und*-questions

			Paired wh-*und*-questions	
		No exhaustive lists	Exhaustive lists	Total
Single wh-questions	No exhaustive lists	5	**0**	5
	Exhaustive lists	12	3	15
	Total	17	3	20

Again, the majority of children (17/20) did not use exhaustivity as a systematic response strategy in paired wh-*und*-questions. Out of the three remaining children all used this response type also with single wh-questions. Thus, taking both analyses together, exhaustivity seems to be mastered before pairing.

4.4. Summary of Experiment 1

While TD children interpreted the exhaustive wh-questions adult-like, children with SLI performed significantly worse than their same age peers. About half of the children with SLI mastered single exhaustive wh-questions without mastering paired exhaustive wh-questions or paired wh-*und*-questions. Out of seven children with SLI who mastered paired wh-questions (and paired wh-*und*-questions) only two (one) respectively responded as though they had not mastered single exhaustive wh-questions. This suggests that exhaustivity in single wh-questions is acquired earlier than in paired wh-questions. That twelve children with SLI gave exhaustive list responses in single but not in paired wh-questions, suggests that additional mechanisms are involved in interpreting multiple wh-questions.

5. Experiment 2

Experiment 1 provided evidence that, unlike children with SLI, five-year-old typically developing German-speaking children interpret single and paired wh-questions targetlike. Experiment 2 served to substantiate this result, using more items per condition (10 rather than 5) and excluding the wh-*alles*-condition, which may have facilitated the exhaustive reading in two of the test versions.

5.1. Participants

We tested 17 5-year-old typically developing children (mean age = 5;5; range = 4;11 to

5;11; SD = 0;4; 8 boys). The children in the TD group all attended regular kindergartens and were reported to not show any signs of language, speech, or hearing impairment. Their inclusion in the TD group was based on their performance in 3 subtests of the SETK 3-5 (Grimm, 2001) designed to reliably detect language impaired children: VS, SG, and MR. All children in the TD group performed at average or above in the at least 2 of 3 subtests, with T- values > 40. (See Appendix B. for the detailed test results).

5.2. Question-with-picture task: design, procedure, materials

This experiment used the same basic design as Experiment 1. The experimenter showed the child a picture, introduced by a short lead-in sentence, and then while the child was looking at the picture asked a wh-question. Each of the participants was tested individually in a quiet room in two sessions; the sessions were about one week apart. In session 1, children received the three subtests of the SETK 3-5. In session 2, children were administered the main experiment. All sessions were video-recorded for later data check against the onsite-coding and for further individual analyses. No response-contingent feedback was given by the experimenter. When the child failed to supply an answer, items were repeated once.

Each subject heard a total of 24 wh-questions, 20 test items and 4 controls. There were 2 conditions, consisting of 10 test items each, which were presented to each participant. Condition one contained single wh-questions and condition two paired wh-questions. As in Experiment 1, all single wh-questions used the wh-pronoun *wer* 'who', and the paired wh-

questions were construed with the fronted subject wh-pronoun *wer* 'who' combined with an accusative or dative wh-pronoun (*wen* 'whom', *was* 'what', *wo* 'where', *mit was* 'with what').

The pictures were developed based on the picture set of Experiment 1. In the single wh-condition, there were between three and six individuals out of which between two and four shared the property being asked about, such as holding a soccer ball. The 2 control items required a non-exhaustive answer, one being a singleton answer and one a rejection of the question, because none of the individuals fulfilled the property. In the paired wh-condition, pictures displayed between two and four individuals sharing the same property such as eating something. The 2 control items required a non-exhaustive pair-list answer: one item depicted two individuals, one of which was engaged in the activity being asked about, and required a single-pair answer, the other depicted two individuals, with none of them being engaged in the activity being asked about. The control items were included to prevent the child from assuming that she always had to respond with more than one pair. The items in the two conditions were presented in a block design. Within a block, test and control items were presented in a random but fixed order.

5.3. Results

The main result was that the 5-year-olds had generally no difficulty understanding exhaustive wh-questions. They performed well and similarly to the five-year-old children in Experiment 1, on both single wh-questions (M = 85.9, SD = 32.8) and paired wh-

questions (M = 84.1, SD = 34.2). No differences was found between the comprehension of single and paired wh-questions, $t(16) = 1.00$, $p = .33$. Performance on single and paired wh-questions was highly correlated (Pearson, two-tailed, $r = .975$, $p < .001$).

Performance on the 10 single wh-items was significantly interrelated (Pearson, two-tailed, all correlations between .595, $p < .05$ and .1.000, $p <. 001$). The 10 paired wh-items are highly interrelated as well (Pearson, two-tailed, all correlations between .595, $p < .05$ and .1.000, $p <. 001$, except for two out of 42 comparisons). These correlations indicate that the children's response pattern is not affected by changes in the number of individuals displayed or in the length of the list answer or PL answer.

Children's performance on the control items varied as expected. While all children gave the correct singleton answer to the single wh-question, only 47.1% correctly rejected the failed presupposition in control item 2. The paired wh-control items were answered correctly by 67.6% of the children, resulting from a failed presupposition in control 1 (70.6%) and an unclear verb for control 2 (64.7%).

REF _Ref231715135 \h Table 11 shows the percentage of correct responses to the two test conditions for each child.

Table SEQ Table * ARABIC 11

Experiment 2. Number of subjects (out of 17) by responses correct for each condition

Percentage of correct responses	Single wh	Paired wh
(10 correct responses possible per subject)		
0% (0/10)	2	2

80% (8/10)	1	2
90% (9/10)	2	3
100% (10/10)	12	10
Mastery (8 or more responses correct)	15	15

As can be inferred from REF _Ref231715135 \h Table 11, only two children performed below criterion in each condition; these were the same children (AM age 5;7; MA age 5;4). REF _Ref231715221 \h Table 12 below illustrates the error types for the two test conditions.

Table SEQ Table * ARABIC 12

Experiment 2. Types of errors by test condition

Condition	Errors Total	Error types
Plural pairs1 PairSubj-listObj-list1 Subj1 ObjOtherSingle wh24/170n.a.n.a.n.a.n.a.23n.a.		
1* paired wh27/17011112-102* Plural list of subjects		

As in Experiment 1, incorrect responses to single wh-questions were predominantly singleton answers. Incorrect responses to paired wh-questions consisted mostly of one-pair answers and one-object answers, with child AM using only the first response type, and child MA only the second type. E.g., to the question *Who is eating* wh*at?* (cf. Figure 2), child AM answered *The boy is eating chocolate*, and child MA replied *A fish.* Defining mastery as before (with 8 out of 10 items correct being significantly different from chance, based on binomial distribution), we also analyzed children's performance across the two conditions. REF _Ref231787878 \h Table 13 shows that two children (AM and MA) did not master both single and paired wh-questions, and that there is no child that masters

paired but not single wh-questions.

Table SEQ Table * ARABIC 13

Experiment 2. Number of children by mastery of single wh-questions and paired wh-questions

			Paired wh-questions	
		No mastery	Mastery	Total
Single wh-questions	No mastery	2	**0**	2
	Mastery	0	15	15
	Total	2	15	17

A Chi-square analysis showed that the distribution of mastery is significantly different from chance, χ^2 (1, N = 17) =$17, $p < .0001$.

In sum, even if the number of characters and the length of the answer list is varied, TD children at the age of 5 consistently give exhaustive answers to single and paired wh-questions. In contrast to Experiment 1, the dominant incorrect response types were one pair and one object answers.

6. General Discussion

In this paper we put forward the hypothesis that single and paired wh-questions should and can be captured in the same approach. Comparing semantic and pragmatic accounts to the interpretation of wh-questions, on simplicity grounds we argued for a semantic account that relates exhaustivity to an inherent property of the question meaning. Pragmatic accounts in our opinion fail to account for the obligatory exhaustivity in multiple wh-questions.

$Under the unified semantic account, we carried out two experiments. Using the question-with-picture task, the two experiments investigated the comprehension of four types of exhaustive wh-questions in 5-year-old TD children and children with SLI. Crucially, a set up was chosen that favoured exhaustive answers, as documented by the adult responses. Experiment 1 compared comprehension of single, paired, wh-*alles*-, and conjoined wh-questions. Experiment 2 tested the first two types with a more extensive design in TD children only. The results of Experiment 1 and 2 indicate that at age 5 typically developing children have mastered the exhaustivity requirement in single and multiple wh-questions, with between 84% and 100% correct responses. Children with SLI had mastered only the *alles*-wh-questions, while in the comprehension of single and paired wh-questions they failed to grasp exhaustivity (13/20) and performed significantly worse than their same-age TD peers.

Confirming the Exhaustivity Acquisition Hypothesis REF _Ref236308529 \r \h (28)a, exhaustivity in multiple wh-questions is recognized later by children with SLI than in single wh-questions. The analysis of the individual children showed that mastery of exhaustivity does not emerge simultaneously across all question types and that in general children who had mastered paired wh-questions had also mastered single wh-questions (cf. REF _Ref231205872 \h Table 7, REF _Ref231205922 \h Table 8). Singleton answers constituted the most frequent error for single wh-questions, corroborating the Semantic Acquisition Hypothesis REF _Ref236308529 \r \h (28)b. Children's response patterns

moreover suggest that once exhaustivity emerges it is applied consistently across all items of one question type. A stage of plural responses does not exist, corroborating the Exhaustivity Acquisition Hypothesis REF _Ref236308529 \r \h (28)c. As older children have not been studied, it remains open, however, whether and at which age children allow an under-exhaustive plural interpretation.

For paired wh-questions, plural PL responses were not found. The most frequent errors in the children with SLI were exhaustive subject lists, object lists, and single pair responses. The same response pattern was found for the conjoined wh-questions. Even though conjoined wh-questions do not require PL answers, contrary to our predictions, the children, like the adults, interpreted them on a par with paired wh-questions. How can these error patterns be explained? Given that cross-linguistically multiple wh-questions can be felicitous in SP contexts, the German learning child should be able to resort to SP answers. However, the SP reponse type constituted only 15% of the errors in the SLI-children (Experiment 1), and was the dominant response pattern for only one TD child (Experiment 2). Even though subject and object lists constituted the majority of incorrect responses for SLI children, it was not the case that children who made use of exhaustive lists in single wh-question always used the exhaustive list strategy, resulting in incorrect exhaustive subject or object lists, in paired wh-questions. This finding suggests that both wh-pronouns were recognized by some children, even though their interpretation then fails. The preference for subject lists in the SLI group is in contrast to two recent studies on typical

acquisition (Heizmann, 2008; Oiry and Roeper, 2009), which found that many children (mostly at age 4) responded to the paired wh-questions with a list of objects. Oiry and Roeper (2009) take their result to suggest that in comprehension children treat the first wh-word as a scope-marker for the second (e.g., *Was glaubt Maria, wen sie besucht?* what thinks Maria, whom she visits? 'What does Maria think whom she'll visit?'). Consequently, children would answer a paired question as *Who is visiting whom?* with a list of objects. On the other hand, the children with SLI who gave exhaustive subject lists may have interpreted the second wh-word as an indefinite, which in its reduced form is homonym with the question word, (e.g., *was* 'what', *(et)was* 'something'). More research is needed to sort out whether language, age, or SLI vs. TD may have caused the different preferences for subject vs. object lists. Independent of the error type found, the great difficulty of children with SLI with paired wh-questions is compatible with Krifka's analysis that according to us$ attributes a greater semantic complexity to paired than to single wh-questions.

Contrary to our prediction, conjoined wh-questions were treated like paired wh-questions and were interpreted incorrectly by the children with SLI. This finding points to the possibility that it is not simply the presence of two wh-pronouns in a clause, but the interpretation of wh-words as pairs that is difficult.

When the exhaustivity marker *alles* was present, all children with SLI provided exhaustive list answers. From this result we conclude that children with SLI are able to

exhaust the domain if the quantificational force is lexically overt, but not if it is covert, as in single, paired and conjoined wh-questions. Fn? One of the reviewers raises the interesting question what this reulst predicts for languages that do not have overt exhaustivity markers. As *alles* does not force plural agreement (Dutch$$),we suggest that plural marking in those languages is not the same.

If our account that the children with SLI have a deficit in semantics is on the right track, then related structures such as universal quantification, which also involve the requirement to exhaust domains, should be difficult for these children as well (for first evidence, cf. Roeper, Pearson, and Strauss, 2005). At the same time, pragmatic phenomena that require fixing underspecified meanings such as pragmatic inferences (e.g., implicatures) should not be difficult for SLI children with deficits in exhaustivity.

In previous work (Roeper et al, 2007), we speculated that the presence of the particle *alles* in German could function as a trigger for exhaustive readings. This way it could be explained why German-speaking children are reported to recognize exhaustivity earlier than English-speaking children (Heizmann, 2008; Roeper et al., 2007). It remains to be seen, however, whether children in languages with overt exhaustivity markers other than German acquire exhaustivity as early as age 4 as well.

Integrating the results of Experiment 1 and 2 and previous research into an acquisition path, the following tentative steps towards mastering exhaustivity can be formulated for typical development:

I. Constant interpretation of all single *wh*-questions(previous research, Experiment 1):

$that is no existential quantifier, but c

II. Overt *alles* is interpreted as exhaustive (Experiment 1)

$no universal quantification involved??

III. Wh-questions are recognized as ambiguous between exhaustive or existential

?stipulation? could also be thattehre is a ophgas ein which all questions are overgeneralizd as exhaustive$

IV. Paired wh-questions are interpreted as requiring exhaustive PL answers (Exp. 2)

V. Plural readings occur in single wh-questions (stipulation!!

Furrther research is called for to substantiate this developmental path. As a first step, in ongoing work, we explore whether the asymmetry between single and paired wh-questions is also found in younger TD children. In addition, using the same test design across typologically different languages, we are currently exploring the degree of language-specific influence on the age of acquisition of exhaustivity and on the intermediate learner grammars (Schulz, 2010).

In conclusion, in this study we argued on a theoretical and intuitive basis against a pragmatic account, which claims that wh-questions are underspecified with respect to exhaustivity. An underspecification account cannot explain the main descriptive findings: PL answers to multiple wh-questions are obligatory exhaustive in the adult grammar. Instead we argued for a semantic account where exhaustivity is rooted in the question

meaning and is independent of pairing. We took the pragmatic account to predict plural answers. However, plural responses were absent in both SLI and TD children across different ages and two languages. In fact, it is interesting that proposals exist that single wh-expressions can have singleton responses, but not paired wh-. Our results are opposite, because children realize exhaustivity on single wh-, before they realize it on multiple wh-. If single and paired wh-questions are unrelated, the null hypothesis is that they show no pattern in acquisition.

Wh- word interpretation obviously involves the semantic notions of exhaustivity and pairing. These notions, in turn, need to be embedded in syntactic representations. We have asserted that the exhaustivity in single wh-expressions and in paired wh-expressions should be essentially the same. This notion is captured by the idea that a feature is present on the wh-word, roughly like the quantifier *every*. Our results show that it appear in a systematic way, first in single wh- environments, and then in paired environments. The property of pairing arises only in multiple wh-environments and calls for a semantics where a mapping function links one wh-word to another. Our results indicate that these properties are distinct because the pairing appears independently and later than exhaustivity.

Many questions remain open. WE do not know if there is an acquisition path within the semantics showing stages in the acquisition of the mapping function. We do not know if pairing can be a pragmatic response independent of the a wh-expression. If, in the experiment above the question would be *What's happening?* instead of *Who is eating what?*, would the children spontaneously provide paired list answers with the PL intonation? And finally, we do not know whether the children know the syntactic contexts in which pairing is not required, such as in conjoined wh-questions *Who ate and what did*

they eat? or island environments (Cheng & Demirdache, $$$$; Krifka, 2001).

Acknowledgements

This research was supported by the (grant no. $). For helpful discussions of exhaustivity in wh-questions we thank Katharina Hartmann, Cecile Meier, Barbara Pearson, Anja Müller, Uli Sauerland, and Malte Zimmermann. We thank the two anonymous reviewers for valuable comments and $$ that helped clarify the ideas presented in this paper. We are grateful to Anja Müller for help with the data analysis and to Paul Abbott for careful comments. Any remaining errors are of course our own.

Appendix A.

Experiment 1. Individual performance on the subtests of the SETK 3-5 for children with TD and with SLI.

		SETK Subtest		
ChildVS (T-value)MR (T-value)PGN (T-value)GW (raw value)SG (T-value)TD DA595855879MO536349557GF726968562NA656361568LE**39***4346667VA657268563ML49**21**49559KO**33**6341666LE535561663NI496655462JO605661454LU597261574JA536643671RA595549464CA535849568AR536343663LA597955768GE4944**35**454LO464955653LA594449746TD Mean				
54.35	57.95	52.65	5.45	63.05
SLI SA**33**44**31**547DE**36**4941**36**MA**34**41**39****35**FR**39**55**37**649AL4343**39****34**SB**223535****37**HE49**37****37****35**VI**33**44**27**3**26**KE**29**41**26****28**JA**39**44**23**443AI**39**48**35****37**JB**333531****36**SO4649**37****33**LP**293527****28**VI4943**23239**TH**36**5343**39**FY**39**46**35****39**PA**29**44**27**340RI**39**44**37****35**JU53**3723**558SLI Mean				
37.45	43.35	32.65	4	37.7

* Bold numbers mark below average performance

Appendix B.

Experiment 2. Individual performance (T-Value) on three subtests of the SETK 3-5 for TD children.

		SETK Subtest	
Child	VS	MR	SG
MA	59	58	54
JO	59	60	54
LE	49	55	46
LU	59	60	50
AM	49	58	58
AN	43	51	55
MA	49	**39***	49
LU	59	**32**	47
LA	59	55	50
MA	65	**27**	57
LL	49	58	45
JO	59	63	**39**
JA			
MA	49	58	59
AL	53	58	50
SO	46	56	
MI	65	58	66
TD Mean	54.44	52.88	51.93

* Bold numbers mark belowe average performance

References

Beck, S., Rullmann, H., 1999. A Flexible Approach to Exhaustivity in Questions. Natural Language Semantics 7, 249–298.

Bošković, Z., 2001. On the Nature of the Syntax-Phonology Interface: Cliticization and Related Phenomena. Amsterdam: Elsevier.

Citko, B., 2008. Wh-questions with coordinated wh-pronouns. Abstract for TABUtag, Groningen.

Costa, J., 2004. Subject Positions and Interfaces: the Case of European Portugese. $City$: Mouton de Gruyter, Band 11.

Dayal, V., 1996. Locality in WH-quantification. Dordrecht: Kluwer.

Dayal, V., 2002. Single-pair versus multiple-pair anwers: wh-in-situ and scope. Linguistic Inquiry, 33, 512-520.

Dayal, V., 2005. Multiple wh questions. In M. Everaert and H. van Riemsdijk, Syntax Companion 3, Case #44. Blackwell Publishers.

Deevy, P., Leonard, L.B., 2004. Comprehension of Wh-questions in children with specific language impairment. Journal of Speech, Language, and Hearing Research 47, 802-815.

De Villiers, J., 1995. Empty categories and complex syntax: The case of wh-questions. In P. Fletcher, B. MacWhinney (Eds.), Handbook of Child Language, 508-540. $PLACE: PUBLSHER$

De Villiers, J., 2003. Defining SLI: A linguistic perspective. In Y. Levy and J. Schaeffer

(Eds.), Language across Populations. Toward a Definition of Specific Language Impairment, pp. 425-447. Mahwah, NJ: Lawrence Erlbaum

Ebbels, S., van der Lely, H.K.J., 2001. Metasyntactic therapy using visual coding for children with severe persistent SLI. International Journal of Language and Communication Disorders 36, Supplement, 345-350.

Friedmann, N., Novogrodsky, R., this volume. Which questions are most difficult to understand?

The comprehension of Wh questions in three subtypes of SLI. $$.Lingua

Grebenyova, L., 2006a. Multiple Interrogatives: Syntax, Semantics, and Learnability. Ph.D. Dissertation, University of Maryland.

Grebenyova, L., 2006b. Multiple interrogatives in child Language. In: D. Bamman, T. Magnistkaia, and C. Zaller (Eds.), BUCLD Proceedings, 225-236. Somerville, MA: Cascadilla Press.

Grimm, H. 2001. SETK 3-5. Sprachentwicklungstest für drei- bis fünfjährige Kinder. Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen. Göttingen: Hogrefe.

Groenendijk, J., Stockhof, M., 1984. Studies on the Semantics of Questions and the Pragmatics of Answers. Ph.D. thesis, University of Amsterdam.

Groenendijk, J., 2008. Inquisitive Semantics. Student Version. Lecture Notes. $$$

Hamblin, C. L., 1973. Questions in Montague English. Foundations of Language 10, 41-53.

Hagstrom, P. 1998. Decomposing Questions. Ph.D. Dissertation, MIT.

Hagstrom, P., 2003. What questions mean. GLOT International, 7/8, 188–201.

Hamblin, C. L., 1973. Questions in Montague English. Foundations of Language, 10, 41-53.

Hartmann, K., Zimmermann, M., 2007. Exhaustivity Particles in Hausa: A Reevaluation of the Particle nee/cee. In: E. Aboh, K. Hartmann and M. Zimmermann (eds.), Focus Strategies in African languages: The Interaction of Focus and Grammar in Niger-Congo and Afro-Asiatic; Berlin: Mouton de Gruyter.241–263.

Heizmann, T., 2007. Children's acquisition of exhaustivity in clefts. In: H. Caunt-Nulton, S. Kulatilake, I.Woo (Eds.), Proceedings of the 31st annual Boston University Conference on Language Development (BUCLD). Cambridge: Cacadilla Press, pp. $$-$$.

Heizmann, T., 2008. Exhaustivität in Spaltkonstruktionen, Fragen und Quantoren bei deutschen und englischen Kindern. Paper presented at Interdisziplinäre Tagung über Sprachentwicklungsstörungen (ISES 5), Mainz, Germany.

Hollebrandse, B., 2003. Distributive qualities of wh-words in Dutch child Language. Paper at Taalkunde in Nederland dag, Utrecht, February$.

Karttunen, L., 1977, Syntax and Semantics of Questions. Linguistics and Philosophy 1, 3-44.

Kazenin, K., 2002. On coordination of wh-phrases in Russian. Ms., Tübingen University and Moscow State University.

Krifka. M., 2001. For a structured meaning account of questions and answers. In C. Fery and W. Sternefeld (Eds.), Audiatur vox sapentia. A Festschrift for Arnim von Stechow. Berlin: Akademie Verlag, 287-319.

Leonard, L., 1998. Children with Specific Language Impairment. Cambridge, MA: MIT Press.

McCloskey, J. 2000. Quantifier Float and Wh-Movement in an Irish English. Linguistic Inquiry 31, 57-84.

Merchant, J., 2001. The syntax of Silence. Sluicing, Islands, and the Theory of Ellipsis. Oxford: Oxford University Press.

Nelken, R., Shan, C.-C., 2004. A logic of interrogation should be internalized in a modal logic for knowledge. Paper presented at SALT 14, Chicago.

Nishigauchi, T., 1999. Some preliminary thoughts on the acquisition of the syntax and semantics of wh-constructions. In: Theoretical and Applied Linguistics at Kobe Shoin 2. Kobe Shoin Institute for Linguistic Sciences. 35-48.

Oiry, M., Roper, T., 2009. How language acquisition reveals minimalist symmetry in the wh-system. In: K. K. Grohmann (Ed.), Explorations of phase theory: Features and arguments. Berlin: Mouton de Gruyter.

Penner, Z., 1994. Ordered Parameter Setting in First Language Acquisition. The Role of Syntactic Bootstrapping and the Triggering Hierarchy in Determining the Developmental Sequence in Early Grammar. A Case Study in the Acquisition of Bernese Swiss German. Habilitation Thesis, University of Bern.

Penner, Z., 1996. From empty to double-filled complementizers. A case study in the acquisition of subordination in Bernese Swiss German. Arbeitspapiere der Fachgruppe Sprachwissenschaft der Universität Konstanz.

Penner, Z., 1999. Screeningverfahren zur Feststellung von Störungen in der Grammatikentwicklung. Edition SZH. Luzern.

Penner, Z., Kölliker Funk, M., 1998. Therapie und Diagnose von Grammatikstörungen: Ein Arbeitsbuch. Luzern: Edition SZH/SPC.

Pérez-Leroux, A.-T., 1993. Empty Categories and the Acquisition of Wh-Movement. Ph.D.Dissertation. University of Massachusetts, Amherst.

Pfau, R., 2008. Weder Wahnsinn noch Willkür: Wh-Fragen ohne Wh-Wort. Handout for the Talk, Graduiertenkolleg Satzarten, Frankfurt University.

Reckling, I., 2005. Der Erwerb von exhaustiven Fragen und Doppelfragen – eine Vergleichsstudie. Unpublished Diploma Thesis, University of Potsdam.

Reis., M., 1992. The category of invariant alles in wh-clauses: On syntactic categories vs. quantifying particles in German. In: R. Tracy (Ed.), Who Climbs the Grammar Tree, pp. 465-492. Tübingen: Niemeyer.

Roeper, T., 2004. Diagnosing language variations: Underlying principles for syntactic assessment. Seminars in Speech and Language 25, 41-55.

Roper, T., 2009. The minimalist microscope: How and where interface principles guide acquisition. In: Chandlee, J., Franchini, M., Lord, S., and Rheiner, G.-M. (Eds.), BUCLD 33 Proceedings. Somerville, MA: Cascadilla Press, 24-48.

Roeper, T., de Villiers, J., 1991. The emergence of bound variable structures. In Papers in the acquisition of WH. Proceedings of the University of Massachusetts Roundtable (UMOP), ed. T. Maxfield and B. Plunkett, 225-266. Amherst MA.

Roeper, T., de Villiers, J., 1993. The emergence of bound variables. In: E. Reuland and W. Abraham (Eds.), Knowledge and Language, 105-139. Boston: Kluwer Academic.

Roeper, T., Pearson, B., Strauss, U., 2005. Disordered language and complex syntax: Quantifier spreading across dialect and impairing status. Paper presented at the Xth International Congress For the Study of Child Language (IASCL), Berlin.

Roeper, T., Schulz, P., Pearson, B. Z., Reckling, I., 2007. From Singleton to Exhaustive: The Acquisition of Wh-. In: Michael Becker and Andrew McKenzie (Eds.), Proceedings of SULA 3, University of Massachusetts Amherst, $$-$$.

Rooy, R. van, 2004. Questions and relevance. In Proceedings of the 2nd CoLog NET-ElsNET Symposium, 96-107.

Schulz, P., Penner, Z. 2002. The emergence of exhaustivity in wh-questions. Evidence from normal and impaired language acquisition in German. Handout, Workshop on "Cooperations in Research on Specific Language Impairment", Groningen.

Schulz, P., Reckling, I., 2005. Deficits in the semantics of single and paired wh-questions as a linguistic marker of SLI? Evidence from German. Paper presented at EUCLDIS, Royaumont.

Schulz, P., Roeper, T., Pearson, B., 2005. The acquisition of the semantics of exhaustive wh-questions from a cross-linguistic perspective. Paper presented at the Xth International Congress For the Study of Child Language (IASCL), Berlin.

Seymour, H. , Roeper, T., and de Villiers, J., 2000. Dialect Sensitive Language Test: Pilot Edition. San Antonio, TX: Psychological Corporation.

Siegmüller, J., Herzog, C., Herrmann, H., 2005. Syntaktische und lexikalische aspeket beim Verstehen von Informationsfragen. Logos Interdisziplinär, 13, 29-35.

Stangen, I., 2008. WER WIE WAS? Die Interpretation von w-Fragen bei fünfjährigen Kindern mit Deutsch als Muttersprache. Unpublished BA. Thesis. University of Frankfurt.

Tracy, R, 1994. Raising questions: Formal and functional aspects of the acquisition of wh-questions in German. In: R. Tracy and E. Lattey (Eds.), How tolerant is universal grammar? Essays on language learnability and language variation, pp. 1-34. Tübingen: Niemeyer

Weissenborn, J., Höhle, B., Penner, Z., 2006. The comprehension of single-member set and multiple-member set questions in 5-year-old German learning children at risk for SLI at the age of 30 months. Results form an interdisciplinary longitudinal study. Paper at COST A 33 Meeting, Lisbon, July.

Zimmermann, M., 2007a. Wer ist so alles zum Vortrag gekommen? – Eine semantische Analyse der quantifizierenden w-Fragepartikeln *so* und *alles* im Deutschen. Handout for the Talk at the GK Satzarten, University of Frankfurt.

Zimmermann, M., 2007b. Quantifying question particles in German: Syntactic effects on interpretation. In E. Puig-Waldmueller (Ed.), Proceedings of Sinn und Bedeutung (SuB) 11, Barcelona, Universitat Pompeu Fabra. 627-641.

Zimmermann, M., in prep.. Quantifying question particles and the non-exhaustiveness of wh-questions, Ms., Potsdam University.

Figure SEQ Figure_ * ARABIC 1

Example picture for a single wh-question (*Who is holding a soccer ball?*)

Figure SEQ Figure_ * ARABIC 2

Example picture for a paired wh-question (*Who is eating what?*)

Figure SEQ Figure_ * ARABIC 3:

Experiment 1. Proportion of corect responses of the TD and the SLI group on the four question types

EMBED Excel.Chart.8 \s

In order to capture the difference between the singleton and the plural answer, in previous work we suggested employing the notions of specificity and definiteness (cf. Roeper et al., 2007).

Note that it is not clear whether these non-exhaustivity markers are simply specified for [+plural] (cf. Zimmermann, 2007b) or whether they call for a subset of few representative members (cf. Beck and Rullmann, 1999).

Flexibility is not an obligatory property of exhaustivity marker, however, as for example in Hausa the focus-sensitive exhaustivity marker has to occur right-adjacent to the sentence-initial focus position (Hartmann, p.c.).

Note that here and in the following we consider the variants with the verb (full term answer?) and without the verb as equivalent (cf. Krifka, 2001).

A reviewer raises the question of how a corrective response with *only* affects this judgment. We argue that indeed in *Only Jane ate a banana* the focus particle takes scope over the whole pair. (cf. Question ($): *Which students failed the exam?* Answer: *Only Mary* (p.15).$)

Since they ask about a particular event, they are also called quiz questions (Krifka, 2001: 16). Dayal (1996) takes REF-questions to involve two existential quantifiers; Dayal (2002) analyzes them as involving choice functions, yielding a SP answer as well.

Costa (2004:173) argues that when a multiple wh-question such as *Quem leu o quê?* 'Who read what?' is answered with SVO, it is not necessarily a complete answer, whereas answers of the form VSO have to be exhaustive.

(i) Who bought what for Max? √PL

(ii) Who wrote a report that Max bought what? #PL

Similarly, Dayal (2002) suggests that intonational differences between REF-questions and regular multiple wh-questions might have gone unnoticed in judging the availability of SP answers.

Expressing this generalization in syntactic terms (cf. Hagstrom, 1998), the originating position of the Q-morpheme (the interrogative morpheme) distinguishes the two interpretations: Merging the Q-morpheme with the lowest wh-phrase results in the PL reading, and merging it with IP results in the SP reading. Some languages then allow the Q-morpheme to select the wh-phrase only, while others allow both selections, wh-phrase and IP (Grebenyova, 2006b). $Grohmann$$

Note that the relation between the two conjuncts is complex. As shown in (i), pro is coindexed with the wh-expression in the first conjunct (cf. Kazenin, 2002, for Russian).

(i) [Who$_i$ [t$_i$ came]] and [why$_k$ [pro$_i$ came t$_k$]]?

Moreover, there seem to be subtle meaning differences between the conjoined wh-question depending on the order of the wh-words, cf. (ii) vs. (ii') that may be related to Kuno's sorting key (cf. Krifka, 2001, 19).

(ii) Why is John painting and what ~~is John painting~~?

(ii') What is John painting and why ~~is John painting~~?

Formalized as follows: $[[?xPx]]^R = \{\lambda v\,[X \in Op(P,v) \cap O(P,w)] \mid w \in W \,\&\, X \in \wp(D)\}$

Note that given this description, under the pragmatic account the notion of default does not apply. The interpretation is said to follow the same rule that yields to different results.

Formalized as follows: w-alles <P,Q> = <P, {x | x ∈ Q & DIV(x) & ¬∃z [z > x & z ∈Q & z ∈ P]}>. The first part takes care of plurality and the second of exhaustivity.

marks infelicitous answers.

We do not take a stand here on whether this approach is superior to the propositional approaches.

See Krifka (2001) for the application to focus and parallels between focus and question-answer pairs, but cf. Pfau (2008) for arguments against the claim that wh-elements are inherently focused.

For emergence of wh-questions in children's speech in German, cf. Tracy (1994).

In previous work, we proposed a formal feature [+variable] to capture the exhaustivity effect (Roeper et al.,

2007). More conservatively, we can say that exhaustivity requires discovering the binding relation between a universal quantifier and a variable. Then, in multiple wh-questions there are two binding relations to be considered.

Alternatively, in propositional semantics, the difficulty could be said to be rooted in the binding relation between both operator and variable, or in not assuming a universal quantifier.

The data presented here have been collected by Ina Reckling in partial fulfilment of her diploma degree (cf. Reckling 2005). The typically developing children were recruited from three day-care centers in Potsdam. The children with SLI were recruited from six speech and language clinics in Potsdam, Berlin, and Brandenburg. We are very grateful to Ina Reckling for her help in collecting and coding the data and to the day-care centers and speech-language therapists for their support.

A one-sample Kolmogorov-Smirnov test showed that for both subject groups the test distribution in all subtests was normal. Thus, T-tests were performed.

The results reported here are a subset of a larger study exploring the relationship between wh-questions and quantifiers. In addition, the data set contained 5 *wh*-questions of the type *Wer malt alles was?* 'Who is all painting what?' and 5 quantifier-questions such as *Reitet jeder Junge auf einem Elefanten?* 'Is every boy riding on an elephant?' In the following, these data will not be considered any further (for details, cf. Reckling, 2005).

Superiority effects were not tested in this Experiment. It would be interesting to explore the difference between *Wer sitzt wo?* and *Wo sitzt wer?*

As one reviewer pointed out, the plural marking of the verb *haben* might insinuate a plural answer. This $CHECK OTHERS$ alle gleich?$ Biasing in this way for plural makes it more convincing for the SLI children's failure..$

The basis for the computer drawings were pictures from SCHUBI Lernmedien AG, which agreed to their use in this experiment.

Performance on the single *wh*-questions in fact differed significantly depending on the test version ($p = .05$). In the two test versions in which single *wh*-questions were presented first, performance was at chance (52% correct), while in the two test versions that presented the single *wh*-questions after the *wh-alles* questions, performance was at ceiling (98%). In the remaining three conditions, the factor test version was not significant.

Note that in the paired wh-control condition, the presupposition of multiple wh-questions, i.e. that they have a list answer (cf. Krifka, 2001), was not fulfilled. The fact that TD children's performance was so high suggests that the context actually qualified as a quiz context. Alternatively, it could be assumed that multiple wh-questions do not carry this presupposition.

Note the SP responses do not refute the generalization proposed in Section 2.3 that Type 1 languages should not allow multiple wh-questions in SP contexts. Crucially, in our experiment the context supplied by the pictures was a multiple pair context.

The data presented here have been collected by Ilse Stangen in partial fulfilment of her Bachelors degree (cf. Stangen, 2008). The children were recruited from three daycare centers in and near Kiel. We are grateful to her for collecting and coding the data and to the daycare centers for their support.

Comparable to the partial movement answers to wh-questions (e.g., *Who$_i$ did the boy ask t_i what$_j$ to buy t_j?*) found by deVilliers et al (1990) and many others, see Strik (2009) for an overview.