Lingua 121 (2011) 383–407

Contents lists available at ScienceDirect

Lingua

journal homepage: www.elsevier.com/locate/lingua

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

Petra Schulz [a,*], Tom Roeper [b,1]

[a] Goethe-University Frankfurt, Institut für Psycholinguistik und Didaktik der deutschen Sprache, Goethe-Universität Frankfurt, Hauspostfach 177, 60629 Frankfurt am Main, Germany
[b] University of Massachusetts, Linguistics Department, South College, University of Massachusetts, Amherst, MA 01030, USA

ARTICLE INFO

Article history:
Received 29 July 2009
Received in revised form 31 March 2010
Accepted 27 September 2010
Available online 22 December 2010

Keywords:
SLI
Wh-question
Comprehension
German
Exhaustivity
Pair-list reading

ABSTRACT

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. A unified semantic approach is suggested that relates exhaustivity to an inherent property of the question meaning. Two question-with-picture experiments explored the comprehension of four wh-question types (single wh-questions with and without the quantifying question particle *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 acquired exhaustivity in single and paired wh-questions. The children with SLI mastered wh-*alles*-questions, but not the other wh-question types. For single wh-questions, the most frequent errors were singleton answers, and for paired and conjoined 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, one which attributes 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.

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. Most research focused on syntactic phenomena such as fronted vs. in situ wh-words and the different structures of object and subject wh-questions, including the differences between *which*- and *who*-questions (see Friedmann and Novogrodsky, 2011; Jakubowicz, 2011; van der Lely et al., 2011). 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 to assess their knowledge of felicity conditions for posing such wh-questions.

* Corresponding author. Tel.: +49 69 798 32561.
E-mail addresses: P.Schulz@em.uni-frankfurt.de (P. Schulz), roeper@linguist.umass.edu (T. Roeper).
[1] Tel.: +1 413 545 6834.

0024-3841/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.lingua.2010.10.005

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

2. Exhaustive wh-questions in semantics and pragmatics

Wh-questions vary across languages in their syntactic and semantic properties (Bošković, 2003; Dayal, 2005; Grohmann, 2003; Hagstrom, 2003). For example, it is well-known that wh-questions may require fronting of all wh-words (Bulgarian), or fronting one wh-word (German, English), or may require the wh-words to be left in situ (Japanese). Moreover, in some languages multiple wh-questions are not allowed (Irish, 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 (Serbo-Croatian, Japanese), in others only PL interpretations are allowed (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[2] (e.g., for which x is it the case that x left?). Thus, by the point of interpretation wh-words – whether one or multiple – must create this operator-variable binding relation (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.

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 a unified 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 (1) to (4) that have to be accounted for in a semantic or pragmatic account.

(1) B gave an exam to 20 students; Peter, Paul, and Mary failed. B's colleague A knows that not all students passed.
A: Who failed the exam?
B: Peter, Paul, and Mary.

(2) A is at a workshop in Germany and is looking for directions to the city centre. 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.

(3) 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.

(4) A: Who is the pope?
B: Joseph Alois Ratzinger.

Example (1) illustrates the mention-all reading, also referred to as 'exhaustive'. Both examples (2) and (3) demonstrate the mention-some reading, also referred to as 'non-exhaustive'. In the following, answer type (2) is called 'singleton' and answer type (3) is called 'plural'.[3] Example (4) presents a special case, as due to world knowledge the exhaustive answer is equivalent to the singleton answer. Note that we are interested in the circumstances under which exhaustive answers are obligatory or not. Therefore, unless stated otherwise in the following discussion the contexts are such that the questioned predicate is satisfied by more than one individual, also called 'more than one answer contexts'; this way the difference between the answers in (1) and (4) is notable. Thus, leaving aside the question of a default reading, all three answer types – exhaustive, singleton, and plural – seem to be available as responses to single wh-questions.

[2] Alternatively, the semantic value of a wh-word is a set of individuals from which a choice function yields a member of that set (see Hagstrom, 1998).
[3] 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 (Roeper et al., 2007).

2.2. (Non)exhaustivity markers

Languages employ various expressions to explicitly mark a wh-question as non-exhaustive or exhaustive (Zimmermann, 2007). First, expressions such as *for example* in English, *so* in German, *zoal* in Dutch (Beck and Rullmann, 1999), and *su* in Hausa (Hartmann and Zimmermann, 2007) mark a non-exhaustive interpretation, more specifically a plural interpretation, as shown in the contrast between (6a) and (6b) as answers to (5).[4]

(5) Wer ist so zur Party gekommen? (German)
'Who for example came to the party?'

(6) a. Mary, Jane, and Sue. (out of 20 guests)
b. #? Mary. (out of 20 guests)[5]

Non-exhaustivity markers thus block a singleton response and trigger a plural answer. Whether *so* and *for example* exclude an exhaustive reading is unclear (Zimmermann, 2007, in preparation).

Second, quantifying question 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, this question is taken to be weakly exhaustive (7a) and (7b) imply that all people who came to the party yesterday have to be identified and are thus incompatible with the singleton answer in (7c). Interestingly, unlike in German, in Dutch the verb can be marked for singular or plural as in (7b) (Bart Hollebrandse, p.c.).

(7) a. Wer alles kam gestern zur Party? (German)
who all came yesterday to_the party
'Who (all) came to the party yesterday?'
b. Wie leest / lezen er allemaal een boek? (Dutch)
who reads_Sg/read_Pl expletive all a book
'Who (all) is/are reading a book?'
c. # Mary.

As illustrated in (8), the position of *alles* is, as in Irish English, variable, with the quantificational effects being the same.[6]

(8) Wer (alles) kam (alles) gestern (alles) zur Party?
'Who (all) came (all) yesterday (all) to the party?'

The semantic status of exhaustivity markers is still under debate. Particles like *alles* and *nèe* have been argued to trigger a conventional implicature; Zimmermann (2007) 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, in preparation). Following Zimmermann (in preparation), 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, in some languages including Italian and Irish, multiple wh-questions are not allowed (Dayal, 2005). In languages that allow multiple wh-questions, they can trigger either a PL or a SP answer (Bošković, 2003; Dayal, 2005; Grohmann, 2002, 2003; Hagstrom, 1998; Krifka, 2001). In English and German, multiple questions, also termed matching questions, generally presuppose that there is more than one answer and hence a SP answer is ruled out (Krifka, 2001). For example, in a situation in which Jane, Tom, and Mary each ate something, the multiple wh-question (9a) requires an exhaustive PL answer as in (9b).[7] Note that here and in the following we consider the variants with the verb (full term answers) and without the verb as equivalents. Also note that we only consider one-to-one pairings, avoiding the question of whether in non-bijective pairings only one element has to be exhausted (see Cheng and Demirdache, 2010; Dayal, 1996). An SP answer is ruled out, as shown in (9b').

[4] Non-exhaustivity markers have been argued to be simply specified for [+plural] (Zimmermann, 2007) or to call for a subset of few representative members (Beck and Rullmann, 1999).
[5] # marks an incorrect answer.
[6] Not all exhaustivity markers exhibit structural flexibility, however. For example in Hausa the focus-sensitive exhaustivity marker has to occur right-adjacent to the sentence-initial focus position (Hartmann, p.c.).
[7] Note that this does not claim that PL answers are only triggered by paired wh-questions; a single question like *What happened?* could in principle be answered with a PL answer as in (9b) as well.

(9) a. Whó ate whát?
b. Jane (ate) a banana, Tom (ate) a sandwich, and Mary (ate) a cookie.
b'. # Jane (ate) a banana.[8]

The presupposition of multiple wh-questions that there is more than one answer seems to be absent in two contexts: with so-called REF-questions and with echo-questions (Comorovski, 1996; Dayal, 1996, 2002; Krifka, 2001; Wachowicz, 1974). REF-questions such as (10) trigger a SP answer.

(10) a. A: Whó hit whóm first?
B: He hit me first.
b. A: Whó killed J.F. Kennedy whén? (Krifka, 2001: 18)
B: Allegedly, Lee Harvey Oswald in 1963.

These questions are about a particular event, and therefore a PL answer is ruled out. In (10a) discourse tells us that only one hitting event took place, and in (10b), world knowledge tells us that the answer is a SP.[9] Second, echo-questions as in (11) also call for an SP answer. Unlike REF-questions, echo-questions refer to a preceding utterance.

(11) A: (almost inaudible): Jonathan ate green spinach.
B: Whó ate whát?

Thus, apart from REF-questions and echo-questions that are felicitous in a SP context, multiple wh-questions in languages like German and English require an exhaustive PL answer in a context that allows for more than one answer.

Interestingly, the answer type 'plural' we noted for single wh-questions (see example (3)), has rarely been discussed with respect to multiple questions.[10] In a world in which it is true that Jane ate a banana, Tom a sandwich, and Mary a cookie, question (9a) cannot be answered with a list of only two of the three pairs, as shown in (12).

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

Note that in the unmarked case the intonation is falling, indicating that the list is closed. Without falling intonation, the speaker seems to indicate that the list is open and may contain other elements, rendering (12) felicitous (see von Stechow and Zimmermann, 1984: Fn 7). A plural PL answer with the unmarked falling intonation contour is only available if the marker *so* is added to one or both wh-expressions as in (13):

(13) Mit wem (so) hat wer denn (so) gesprochen?
'With whom (for example) did who (for example) talk?'

As illustrated in (13), *so* has the effect of cancelling the exhaustivity requirement in multiple wh-questions, which is similar to its semantic contribution in single wh-questions, suggesting a unified 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, as shown in the contrast between (14) and (15) (Hagstrom, 2003):

(14) Who bought what for Max? $\surd$PL

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

Second, the distribution of SP and PL readings is subject to cross-linguistic variation. As mentioned above, in English and German, as well as in Bulgarian, Romanian, Brazilian Portuguese, and Russian, multiple wh-questions like (9a) are not felicitous in SP contexts, and hence do not allow SP answers like (9b') (Bošković, 2003; Grohmann, 2002, 2003; Hagstrom, 1998). However, the SP reading is freely available in Serbo-Croation, Japanese (Bošković, 2003), and Malayalam (Grebenyova, 2006a). In other words, questions like *Who ate what?* are felicitous in both PL and SP contexts. A summary of the occurrence

[8] A reviewer raises the question of how a corrective response with *only* affects this judgment. Singleton answers with corrective *only* are possible for wh-questions with plural wh-pronouns (i) and with *alles* (ii):
(i) Q: Which students failed the exam? A: Only Jane.
(ii) Q: Wer kam alles? 'Who all came?' A: Nur Jane. 'Only Jane.'
We argue that *Only Jane ate a banana* is a possible answer to (9a) and in that case *only* takes scope over the pair (see also Zimmermann, in preparation, for a discussion of *only*).

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

[10] But see Costa (2004) for European Portuguese. He 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 (Costa, 2004: 173).

Table 1
Availability of bare multiple wh-questions across languages.

Context	Type 1 (*SP)	Type 2 (SP&PL)	Type 3 (*PL)	Type 4 (*SP&*PL)
	English, German, Russian, Bulgarian, Brazilian Portuguese, Romanian	Serbo-Croatian, Japanese, Malayalam	?	Italian, Irish
Pair list (PL)	✓	✓	*	*
Single pair (SP)	*	✓	✓	*
Quiz context (Q-SP)	✓	✓ (by inference from the SP context)	✓ (by inference from the SP context)	*

of multiple wh-questions across languages is given in Table 1. Interestingly, to our knowledge there are no languages that allow only SP readings but not PL readings for multiple wh-questions (Type 3). As suggested by a reviewer, a Type 4 can be added with languages including Irish and Italian that do not allow multiple wh-questions in either context.[11] Taking into account effects of superiority and D-linking of wh-pronouns, Grohmann (2003) suggests what he calls the Hagstrom-Bošković approach to multiple questions. According to this syntactic-semantic approach, wh-questions can be classified in a more refined way than depicted in Table 1 regarding the availability of PL and SP interpretations across languages depending on zero-, single or multiple wh-movement. These structures are beyond the scope of the present paper.

A caveat is in order regarding the findings summarized in Table 1: Whether the REF-readings in the quiz context in Type 1 languages (e.g., Krifka, 2001, for German) may in fact be equivalent to the unmarked SP context in Type 2 languages, has to be left open. It would result in merging Type 1 and Type 2 languages.[12] 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.[13] In other words, Type 1 is taken to be the default option.

2.4. Conjoined multiple wh-questions

Conjoined wh-questions constitute a different type of multiple wh-questions, illustrated in (16). According to Krifka (2001), they do not presuppose a PL answer and hence are felicitous in an SP context.

(16) a. A: Whó left and whén?
B: Máry left, at fóur.
b. A: Wánn hast du studiert und wó?
When have you studied and where?
'When did you study, and where?'
B: In den Achtzigern, in Heidelberg.
'In the Eighties, in Heidelberg.'

Note that the wh-words can be conjoined and fronted by themselves as in (17a). As a response to (17a) an SP answer like (16b-B) is preferred, while in (17b) a PL answer is favoured (see Citko, 2008). In this regard (17a), is interpreted parallel to the conjoined wh-question in (16b).

(17) 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?'

[11] According to Adriana Belletti (p.c.), even in quiz-contexts multiple wh-questions are not allowed in Italian.
[12] 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.
[13] Expressing this generalization in syntactic terms, the originating position of the interrogative Q-morpheme (high vs. low) has been argued to distinguish the two interpretations: If [Q] moves from some clause-initial position to C^0, PL readings are generated; if [Q] is generated above both wh-pronouns and moves to C^0, SP readings occur (see Bošković, 1998; Hagstrom, 1998). Some languages then seem to allow only the first option, while others allow both options.

Conjoined questions involve sluicing as shown in (18a), where the ellipsis site is marked by strikethrough (Merchant, 2001). The structure underlying the second conjunct is assumed to be the one in (18b), where governing an empty category in IP is allowed only for C^0 marked as a question via the feature [+Q].

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

In Russian it is impossible to strand a wh-argument (Kazenin, 2002). In contrast, this argument/adjunct asymmetry does not seem to hold for German: Stranded wh-expressions may not only be adjuncts as in (19a) but also arguments as in (19b). Both structures were tested in Experiment 1. Note that in certain contexts the fronted wh-expression in the first conjunct may also be an adjunct as in (20).

(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?'

(20) Warum fährt Maria weg und mit wem?
why drives Maria away and with whom
'Where is Maria going and with whom?'

Grammaticality in conjoined wh-questions – at least in German – thus seems to depend not on the ban of stranding a wh-argument but rather on the grammaticality of the first conjunct. More precisely, all obligatory theta-roles of the verb of the first conjunct have to be filled in order for the conjoined question to be grammatical.[14] Regarding their semantics, conjoined questions are answered one question at a time, as shown in (21) (Krifka, 2001: 22).

(21) 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.

In contrast to multiple wh-questions in Type 1 languages, an SP context is felicitous for a conjoined question such as (21). To our knowledge, PL contexts for conjoined questions have not yet been considered. Given the analysis in (21) for SP contexts, in a PL context (Mary came at four, John at five, and Sue at six), the conjoined question (21a) should not be answered with a single conjoined answer like (21d). 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 (21b-d), such as *Mary, at four; John, at five; and Sue, at six*.

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. Rather than asking when a wh-question is felicitous we focused on the types of answers that 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 generally infelicitous in contexts that only require one answer in Type 1 languages. Table 2 summarizes the types of answers that are available in German for the various types of wh-questions considered, given that the context allows for more than one answer.

[14] 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 (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-questions depending on the order of the wh-words (see (ii) vs. (ii')) that we speculate may be related to Kuno's sorting key (see Krifka, 2001: 19).
(ii) Why is John painting and what ~~is John painting~~?
(ii') What is John painting and why ~~is John painting~~?

Table 2
Felicity of answer types across different types of wh-questions (in contexts with more than one answer).

Question type	Answer 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 the exhaustive and non-exhaustive answer types, depicted in Table 2, be derived? The various proposals center around the questions of whether the distinction between mention-some and mention-all is a semantic or a pragmatic affair (Dayal, 2005), and whether wh-questions are ambiguous or underspecified (van Rooij, 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, a unified semantic one, and will argue that the acquisition data is compatible with this semantic approach.[15]

According to the pragmatic account, the actual interpretation of a wh-question is underspecified by its conventional meaning (van Rooij, 2004; Zimmermann, 2007; in preparation). van Rooij (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, would denote (one of the) optimal values of P in w.[16] 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.[17] In this line of reasoning, it is assumed that wh-questions are not semantically specified as inherently exhaustive and that the mention-all reading may come about as a result of a pragmatic interpretation strategy. According to Zimmermann (in preparation), *alles* then directly adds exhaustivity to the core propositional content of the question and its corresponding answers, the exhaustivity that is not present in single wh-questions without *alles*.[18] 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 generally answered with an exhaustive PL, assuming the standard falling intonation.

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 means to mark (non)exhaustivity. Besides quantifying question particles (see Section 2.2), in languages like English and German 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 different structures should exist, even in the unmarked case of *who*-questions. This is exactly the approach proposed by Nelken and Shan (2004), which we adopted in the past (Roeper et al., 2007). 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 (22), exhaustive wh-questions such as (23) 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.

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

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

[15] We leave open at this point whether this semantic account could be related to something ultimately syntactic in nature, based on the syntactic operations involved to compute the relevant interpretations, as mentioned by one reviewer.
[16] Formalized as follows: $[[?xPx]]^R = \{\lambda v[X \in Op(P, v) \cap O(P, w)] | w \in W \& X \in \wp(D)\}$.
[17] A reviewer asks whether there is a default under the pragmatic account as well. As the interpretation of the operator yields different results depending on what is most optimal in w, in our view the notion of a default does not apply.
[18] 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.

As an alternative to this propositional approach, in a structured meaning approach question meanings are understood as functions that when applied to the meaning of the answer yield a proposition. Krifka (2001) explicitly addresses the meaning of multiple and conjoined wh-questions and in the following we adopt his structured meaning approach, adding a (non) exhaustivity operator to derive the intended readings. Wh-expressions such as *wer* 'who' denote atomic individuals and plural individuals, marked by the plurality operator * (Link, 1983), but not things or animals, as illustrated in (24) (see also Zimmermann, in preparation):

(24) $[[\text{who}]] = \{x|x \in *\text{PERSON}\}$
$= \{\text{Mary}, \text{Jane}, \text{Sue}, \text{Mary} + \text{Jane}, \text{Jane} + \text{Sue}, \text{Mary} + \text{Jane} + \text{Sue}, \ldots\}$

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 (25).[19]

(25) $[[\text{Who left?}]] = <\lambda x.x \text{ left}, \{x|x \in *\text{PERSON}\}>$

Informally speaking, in our semantic account 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 Nishigauchi (1999) suggested that the wh-element in questions contains a variable that is quantificational in nature, at least in languages like English. Groenendijk and Stockhof (1984) derive the exhaustiveness by means of a silent EXH-operator, which is inserted into the answers's logical form. Additional support for the assumption that wh-pronouns host a hidden universal quantifier 'every' comes from an 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?* (requiring an exhaustive list of subjects). Heizmann found that children who fail to exhibit exhaustivity in wh-questions also fail to assign target-like scope to universally quantified statements.

Notably, the notion that wh-questions are ambiguous is compatible with both the view that the exhaustive reading and the view that 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; Karttunen, 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 question meanings.

The answer possibilities for multiple wh-questions, which differ according to the given PL or SP context, are in the semantic account 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 as answers to multiple wh-questions in PL contexts, 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 (26) (that is asking for more than one thing at the time) is transformed by the operators in (27) to a question that asks for a mapping procedure (Krifka, 2001: 23):

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

(27) 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 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* (28).

(28) 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 (26) to (28) 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*.

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

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 in Type 1 languages like German. Plural PL answers and SP answers are excluded. Only in an SP context such as a quiz situation is an SP answer felicitous. Addition of an exhaustivity marker like *alles* in German 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. In our view, under a pragmatic account this absence of a plural PL answer in paired wh-questions has to be explained by stipulating different mechanisms for single and paired wh-questions.

Therefore, in the following, we assume a unified 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 addressed next is whether the semantic account is compatible with the acquisition data.

3. Acquisition of exhaustivity in wh-questions

3.1. Comprehension of singleton wh-questions

Comprehension of singleton wh-questions, i.e. wh-questions in one answer contexts, has been examined in several studies across languages, focusing on subject and object questions. For typical development, evidence is mixed with regard to the difficulty of these question-types. Some studies reported no difference between subject and object questions (Deevy and Leonard, 2004; Stromswold, 1995), while other studies found differences in performance between subject and object questions (De Vincenzi et al., 1999; Jakubowicz and Gutierrez, 2007). Similarly, Siegmüller et al. (2005) reported a general advantage of interpreting subject over object questions and of argument over adjunct questions in German. 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 et al. (2009) reported an advantage of *which*-subject questions over *which*-object questions for Hebrew, raising the possibility that the type of wh-pronoun plays a role in the object/subject asymmetry.

Little research focused on the comprehension of wh-questions in children with SLI (see Friedmann and Novogrodsky, 2011, 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 (Ebbels and van der Lely, 2001; Friedmann and Novogrodsky, 2011; Siegmüller et al., 2005). Siegmüller et al. (2005) found that the German-speaking SLI group performed significantly better on subject than on object questions and, compared to their same age TD peers, performed significantly worse in all question types (subject, object, adjunct) (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 et al. (2006) reported that at age 5 children in the low language ability group performed worse on singleton wh-questions than those in 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 (de Villiers, 1995), with the age of mastery differing across languages, and possibly also according to argument-type. Extending the research by Roeper and de Villiers (1991, 1993), and Pérez-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 variant of the question-after-story task (cf. Roeper and de Villiers, 1991), Schulz and Penner (2002) found that 6- and 7-year-old German-speaking TD children performed almost like the adult controls (age 6: 85%; age 7: 84%, adults: 98%). The method used by Schulz and Penner (2002), which we will refer to as the question-with-picture task, was the origin of the two experiments presented in this paper. It worked 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. In the eight test trials, between two and five characters exhibited the property, while in the four control trials none or one of the characters exhibited the property being asked about.

A cross-linguistic study of the acquisition of exhaustive wh-questions in English and German (Roeper et al., 2007) 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 (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 ages 4 and 5 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 plural marking on the verb (*Wie lezen 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 at age 6 and age 7 less than half of the children with SLI have mastered exhaustivity (6/16 and 7/16 children, respectively), compared to 81% of the TD children at both age 6 and age 7 (13/16 children each). The majority of errors in both SLI groups consisted of singleton answers. Weissenborn et al. (2006) tested 5-year-old children on single exhaustive wh-questions. Children in the low language group performed like average language-performers (see Section 3.1), with 64% correct responses. Only 10% of the errors were plural answers.

3.3. Comprehension of multiple wh-questions

Even though multiple wh-questions are infrequent in the parental input,[20] paired wh-questions have been noted to occur in children's speech around age 3. In a cross-linguistic study of the production of paired wh-questions in PL and SP contexts, Grebenyova (2006a,b) 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, though to a lesser degree than the adults tested (14% vs. 44%, Grebenyova, 2006a: 183).

Only a few studies have looked into the comprehension of multiple wh-questions, and all of them 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 exhaustive paired lists, 4- to 6-year-olds responded with exhaustive paired lists in 78% of the cases and younger children in only 32% of the cases. This developmental pattern was confirmed for German by Heizmann (2008). The number of PL responses to paired wh-questions increased with age, with about 20% PL answers at age 3 and 90% PL answers at age 5. Interestingly, in English for the same task mastery of multiple wh-questions was found later, with only 60% PL responses at age 7 (Heizmann, 2008). Both German- and English-speaking children's incorrect responses were mostly exhaustive object lists. Based on data from a standardized test (DSLT, Seymour et al., 2000), Roeper (2004) reported that children with language impairment had persistent difficulty with the comprehension of multiple wh-questions and performed significantly below their same-age TD 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 typical 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. For example, in *Who is eating what?* the set of subjects and the set of objects being eaten have to be related. However, as only group data are reported, 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 of the acquisition of both question types remains open for SLI as well, especially given 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.

3.5. Research questions and hypotheses

The main goal of the two experiments 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 predictions for acquisition can be derived from the unified semantic account of wh-questions? Recall that under the semantic view, exhaustivity is a feature represented in the structure of both single and multiple wh-questions. Overt

[20] Grebenyova (2006a: 160f) reports that searches in the CHILDES corpora yielded one paired wh-question compared to 697 single wh-questions for Russian and three paired wh-questions compared to 5000 single wh-questions for English.

markers such as *for example* relax the exhaustivity requirement. Extending Krifka's (2001) account, exhaustivity is formalized as a universal quantifier exhausting the question domain, which varies depending on the type of wh-pronoun, for example PERSON for *who* or THING for *what*.[21] 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 crucially rest on the discovery of the relation between universal quantifier and the function domain. In contrast, plural answers hinge on encountering contexts in which a *for example*-answer is called for. From this semantic line of reasoning, some specific predictions for acquisition can be formulated, summarized in (29) as the Semantic Exhaustivity Acquisition Hypothesis. As exhaustivity in single and multiple wh-questions has largely been considered independently, these considerations are necessarily speculative.

(29) Semantic Exhaustivity Acquisition Hypothesis

1. For complexity reasons, exhaustivity in multiple wh-questions is recognized later than in single wh-questions.
2. Before mastery of exhaustivity, wh-pronouns are consistently interpreted as a constant, resulting in singleton responses.
3. Plural responses are not an intermediate acquisition step; they are acquired only after exhaustivity is mastered.

Note that in (29-2), initially the wh-pronoun is represented as a constant and not as a variable involving an existential quantifier, because the latter representation would also allow for plural answers, which are not derivable when a constant is present.

In addition, under both semantic and pragmatic accounts, wh-*alles*-questions should be acquired early, as *alles* carries its own exhaustivity feature. Moreover, 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, in the following we assume that children with SLI in principle follow the same acquisition path as TD children (for discussion of delay vs. deviation, see de Villiers, 2003; Leonard, 1998; for discussion of the nature of SLI see Marinis, 2011). Consequently, the Semantic Exhaustivity Acquisition Hypothesis is assumed to hold for both TD and SLI. The specific difficulty in children with SLI is stated in (30), predicting a delay in acquisition compared to TD children (modifying a previous version in Roeper, 2009).

(30) Missing Quantifier Hypothesis

Children with SLI do not recognize that the question domain has to be exhausted.[22]

We leave open at this point whether children with SLI will arrive at the same target grammar as TD children, strengthening the assumption of a delay, or whether children with SLI show acquisition patterns that are qualitatively different from TD children, suggesting an additional deficit (see Schulz, 2010b, for a deficit account). As research on difficulties with semantic/pragmatic phenomena is just beginning to emerge, we also 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, 2010b).

4. Experiment 1

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 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 favouring non-exhaustive answers in general. As a starting point, only simple wh-expressions were tested.

4.1. Participants

We tested 20 5-year-old typically developing children (mean age = 5;8; range = 5;0–5;11; SD = 3,5 months; 14 boys) and 20 5-year-old children with SLI (mean age = 5;4, range = 5;0–5;10, SD = 3,1 months; 10 boys).[23] In addition, 20 adults served as a control group (age range = 25–57). They were tested individually, following the same procedure and using the same material as the children, except for the SETK 3-5 (Grimm, 2001) and 2 pretests.

[21] 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. In multiple wh-questions there are two binding relations to be considered.

[22] 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.

[23] The data presented here were collected by Ina Reckling in partial fulfilment of her degree requirements (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.

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, which contains five subtests aimed at diagnosing children with SLI. Two subtests assess morpho-syntactic abilities: Subtest VS assesses comprehension of sentences in varying complexity (act-out-task), and Subtest MR tests knowledge of plural marking (elicited production). Three subtests assess memory related abilities, using elicited imitation: Working memory is assessed via repetition of non-words in Subtest PGN, of real words in GW, and of sentences with 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 (see 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. 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 with 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: $t(38) = 10.53$, $p < .001$.[24]

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 about 4 weeks apart. In session 1, children received two subtests of the SETK 3-5 (VS, SG), a wh pretest, 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.[25] There were 4 conditions, each comprising 5 test items, which were presented to each participant. Condition 1 contained a single wh-pronoun (henceforth single wh-question), condition 2 contained a single wh-pronoun and the lexical exhaustivity marker *alles* (henceforth single wh-*alles*-question), condition 3 contained two wh-pronouns (henceforth paired wh-question), and condition 4 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.[26]

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 (31) and Fig. 1.

(31) Guck mal, was ist denn hier los?
'Look, what is happening here?'
a. Wer hat einen Fußball?
who has asoccer ball
'Who is holding a soccer ball?'

[24] 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.

[25] 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, this data will not be considered any further (for details, see Reckling, 2005).

[26] Note that superiority effects were not tested in this experiment, so the order was always: Wh_1-subject wh_2-non-subject.

Fig. 1. Example picture for a single wh-question (*Who is holding a soccer ball?*).

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

For paired wh-questions, a sentence describing the scene was added, without referring to single individuals or specific activities. Typical test items for paired wh-questions and paired wh-*und*-questions are given in (32) and Fig. 2.

(32) Die haben Hunger. 'They are hungry.'[27]
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)?'

Each picture displayed several family members, which were introduced in an initial picture as a family, naming them *mother, father, grandfather, grandmother, boy, girl* to minimize memory effects.[28] In the single wh- and the single wh-*alles*-condition, there were always six individuals of which between two and five shared the property in question, such as holding a soccer ball. This variation ensured that children could not develop guessing strategies, such as listing all individuals in the picture or

[27] As one reviewer pointed out, the plural marking of the verb *haben* might insinuate a plural answer. Other lead-in sentences contained collective nouns and plurals maybe facilitating plural answers as well (*The family has pets. Who is feeding whom*?). Note that this way, singleton answers given by the participants cannot be simply attributed to a context favouring singleton answers.
[28] The basis for the computer drawings were pictures from SCHUBI Lernmedien AG, which agreed to their use in this experiment.

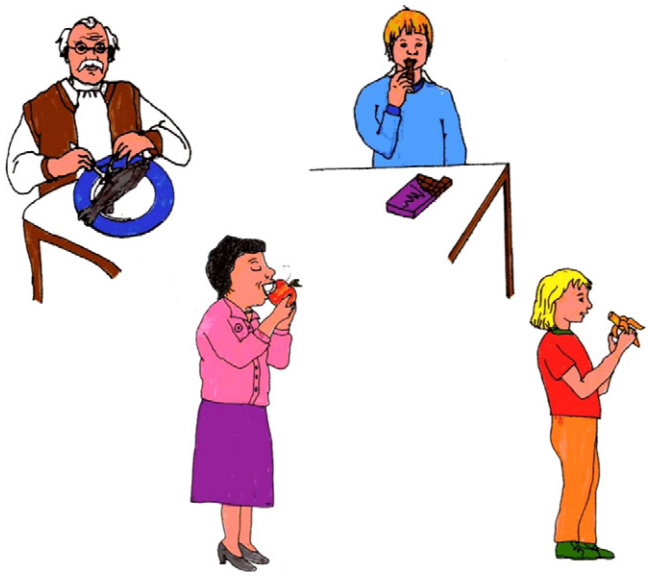

Fig. 2. Example picture for a paired wh-question (*Who is eating what?*).

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 in question, and required an 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 (*drink, eat, read, sit,* etc.).

Table 3 summarizes the coding for the different types of correct responses (for a detailed analysis of the error types, see Tables 4 and 5). Note that the coding for the responses to wh- and wh-*alles*-questions was the same.

Importantly, correct answers to paired wh-*und*-questions are of two different types: responding with two separate for subjects and objects or responding with a list of pairs. We chose this coding option for two reasons. First, a detailed lists

Table 3
Types of correct responses to the test items.

Condition	Response type	Example
Wh(-alles) (*Who (all) has a soccer ball?*)	Listing the subjects with VPs	*The dad has a soccer ball, the child has a soccer ball, and the boy has a soccer ball*
	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?*)	Listing pairs with verb	*The girl is eating a banana, the boy chocolate, the grandma an apple, the grandpa a fish*
	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*
	Listing pairs with verb	*Mom reads the newspaper, and Grandpa reads a book*
	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)

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 confirmed by the adults' preference for PL answers in the current experiment (93%).

The items in the four conditions were presented in a block design to minimize carry over effects of the 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 the order of the conditions was varied while the beginning with single wh-questions remained constant.

4.2.2. 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 (*Reto und sein Vater machen einen Kuchen* 'Reto and his father are making a cake'), followed by a wh-question (*Mit wem macht Reto einen Kuchen?* 'With whom is Reto making a cake?'). This pretest established that both the TD and the SLI group were 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 (soccer ball, chocolate, etc.). 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.

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. Note that this result also indicates that the children with SLI did not exhibit general difficulties with wh-movement structures.

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. Analyzing 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. Thus, the children with SLI did not exhibit a lexical deficit that could negatively influence their performance on the main experiment.

4.3.2. Main experiment

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

An analysis of variance (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, $F(3, 32) = 1.73$, $p = .181$. In the following, data was therefore 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 Fig. 3. 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),[29] 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*-

[29] 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.

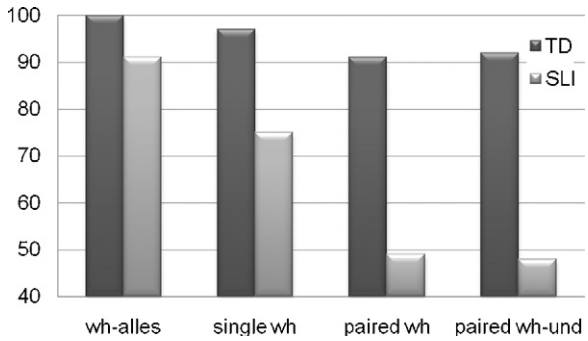

Fig. 3. Experiment 1. Proportion of correct responses of the TD and the SLI group for the four question types.

questions, where TD children's performance was at ceiling. Children with SLI performed significantly poorer than the TD group in the remaining three 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 an SP response.[30]

Error analysis. Children's incorrect responses were then grouped according to error type to more closely examine the source of their 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 Table 4, most of SLI children's errors in single wh-questions were singleton responses.

Table 4
Experiment 1. Types and number of errors for single wh-questions by subject group (number of the participants out of 20 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	25 of 100	2 (2)	23 (5)
	wh-alles	9 of 100	2 (2)	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 Table 5. Notably, plural pair responses effectively 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. Exhaustive list of subjects are used by 12 of the children with SLI. Other recurrent errors among the children with SLI are answers that consist of an exhaustive list of objects, used by 6 children, and of a single pair, used by 8 children.[31]

[30] Note that in the paired wh-control condition, the presupposition of multiple wh-questions, i.e. that they have a list answer (Section 2.3), 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.

[31] Note that 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.

Table 5
Experiment 1. Types and number of errors for paired wh-questions by subject group.

Group	Condition	Errors total	Error types						
			Plural pairs	Single pair	Subj-list	Obj-list	1 Subj	1 Obj	Other
TD	Paired wh	9/100	–	–	3	3	–	–	3[a]
	Paired wh-*und*	8/100	1	–	4	–	2	–	1
	Total	17/200	1	–	7	3	2	–	4
SLI	Paired wh	51/100	–	8	23	12	–	2	6
	Paired wh-*und*	52/100	–	8	19	9	7	6	3
	Total	103/200	–	16	42	21	7	8	9

[a] Saying *alle/jeder*.

Table 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

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. Table 6 shows the percentage of correct responses to the four test conditions for each child in the two subject groups.

Mastery of a wh-question-type was defined strictly 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 (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. 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.

A variety of individual error patterns was observed for the paired wh-questions and the paired wh-*und*-questions. Most notably, the single pair-error, being observed 8 times each across children, was not the favoured strategy of any of the children. 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. 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.

Subsequently, 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. Tables 7 and 8

Table 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 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

Table 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

Table 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

illustrate the number of children who mastered single wh-questions and one of the other question types. Note that *wh-alles*-questions were excluded from the analysis as SLI children's performance was near ceiling.

As can be seen in 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. 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 investigation of the response type 'exhaustive list' across conditions. SLI children who failed paired wh-questions by systematically responding with exhaustive lists of individuals (only subjects, only objects, or both) could make use of this pattern in their responses to the single wh-questions as well. Tables 9 and 10 illustrate the number of children who exhibit exhaustivity in lists across question types. The response pattern 'exhaustivity' is attested if at least 4 out of 5 responses fall into this category. Note that the majority of children (16/20) did not use exhaustivity in lists 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 (Child JB).

Again, the majority of children (17/20) did not incorrectly use exhaustivity in lists as a systematic response strategy in paired wh-*und*-questions. The three remaining children also used this response type with single wh-questions. Thus, taking both analyses together, exhaustivity seems to be mastered independently and 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 TD peers, except for wh-*alles*-questions. About half of the children with SLI mastered single exhaustive wh-questions without mastering paired exhaustive wh-questions or paired wh-*und*-questions. Of the 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 consistently provided 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 target-like. 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;11–5;11; SD = 0;4; 8 boys).[32] 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 designed to reliably detect language impaired children: VS, SG, and MR. All children in the TD group performed at average or above in 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 1 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 accusative or dative wh-pronouns (*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 two single 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 two paired 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 TD 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 were 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, $p < .001$), as was performance on the 10 paired wh-items (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 was not affected by changes in the number of individuals displayed or in the length of the list or PL answer, respectively.

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 the second control item. The paired wh-control items were answered correctly by 67.6% of the children, resulting from a failed presupposition in control item 1 (70.6%) and an unclear depiction of the verb (*open*) for control item 2 (64.7%). Table 11 shows the percentage of correct responses to the two test conditions for each child.

[32] The data presented here was collected by Ilse Stangen in partial fulfilment of her Bachelors degree requirements (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.

Table 11
Experiment 2. Number of subjects (out of 17) by responses correct for each condition.

Percentage of correct responses (10 correct responses possible per subject)		Single wh	Paired wh
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

Table 12
Experiment 2. Types of errors by test condition.

Condition	Errors total	Error types						
		Plural pairs	1 Pair	Subj-list	Obj-list	1 Subj	1 Obj	Other
Single wh	24/170	n.a.	n.a.	n.a.	n.a.	23	n.a.	1[a]
paired wh	27/170	1	11	1	2	–	10	2

[a] Plural list of subjects.

Table 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

Mastery was defined as consistent use of exhaustive answers within each condition (single, paired) and set at 80% correct. As can be inferred from Table 11, only two children performed below criterion in each condition; these were the same children (Child AM, aged 5;7; Child MA, aged 5;4). Table 12 below illustrates the error types for the two test conditions.

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. For example, to the question *Who is eating what?* (Fig. 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), we also analyzed children's performance across the two conditions. 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 mastered paired but not single wh-questions. A Chi-square analysis showed that the distribution of mastery is significantly different from chance, χ^2 (1, $N = 17$) = 17, $p < .0001$.

In summary, even if the number of characters and the length of the answer list is varied, TD children at the age of 5 predominantly 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

This paper investigated how exhaustivity in single and multiple wh-questions is acquired in German-speaking children with SLI. First, 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, we argued for a unified semantic account that relates exhaustivity to an inherent property of the question meaning in single and multiple wh-questions. Pragmatic accounts in our opinion fail to explain the obligatory exhaustivity and the absence of plural answers for multiple wh-questions.

Second, we carried out two question-with-picture experiments, which investigated the comprehension of different types of exhaustive wh-questions in 5-year-old TD children and children with SLI, and explored whether the unified semantic account is compatible with the acquisition data. For both experiments, 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 correct responses ranging from 84% to 100%. Changes in the number of individuals depicted or in size of the answer set did not change TD children's response pattern. In contrast, children with SLI have mastered only the wh-*alles*-questions. In the comprehension of single and paired wh-questions the SLI group performed significantly worse than the same-age TD group. 65% of the children with SLI failed to grasp exhaustivity in paired and conjoined wh-questions and 25% did not master exhaustivity in single wh-questions.

How do these results relate to the predictions postulated by the Semantic Exhaustivity Acquisition Hypothesis (29)? The results from Experiment 1 show that exhaustivity in multiple wh-questions is recognized later by children with SLI than exhaustivity in single wh-questions, confirming prediction (29-1). An analysis at the subject level indicated 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 (Tables 7 and 8). In addition, although the 5-year-old TD children showed in general mastery of exhaustivity, there was one child each who had mastered single but not paired wh-(*und*-) questions in Experiment 1 (Table 6). In Experiment 2, there were no TD children who showed mastery of paired before mastery of single wh-questions (Table 13).

Prediction (29-2) states that before they master exhaustivity children interpret wh-pronouns as constants, resulting in singleton answers. Singleton answers constituted indeed the most frequent error for single wh-questions in the SLI group (Table 4) and in the TD group (Exp. 2, Table 12), corroborating (29-2). TD children's response patterns moreover suggest that once exhaustivity emerges it is used consistently across single wh-questions, thus performance is either at 0% correct or at 80% or above (Tables 6 and 11). In line with previous findings (Roeper et al., 2007; Schulz and Penner, 2002; Weissenborn et al., 2006), plural responses were not existent in either TD or SLI children (Tables 4 and 12). Thus, an intermediate stage of plural responses does not exist for single wh-questions, corroborating prediction (29-3). As older children have not been studied, it remains open, however, whether and at which age children allow an under-exhaustive plural interpretation as in the bathroom example (3).

When the exhaustivity marker *alles* was present, all children with SLI provided exhaustive list answers, as predicted under semantic and pragmatic accounts. 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 wh-questions.[33]

Turning to the response types for multiple wh-questions, plural PL responses were not found in either TD or SLI children (Tables 5 and 12), providing further support for (29-3). The most frequent errors in the children with SLI were exhaustive subject lists, object lists, and single pair responses. Single pair responses could be formalized as treating both subject and object as a constant, thus confirming (29-2). Subject and object lists could be represented as exhausting over one wh-pronoun only. That would mean that the second wh-pronoun is not interpreted and a constant is not present, compatible with prediction (29-2). Although conjoined wh-questions do not require PL answers, possibly due to task effects children and adults interpreted them on a par with paired wh-questions. Thus, the same response patterns were found for both wh-question types. How can the error patterns observed in both multiple wh-questions then be explained? Given that cross-linguistically multiple wh-questions can be felicitous in SP contexts, the German-learning child could be assumed to resort to SP answers also in PL contexts. However, the SP response type constituted only 16% of the errors in the SLI-children (Experiment 1), and was the dominant response pattern for only one TD child (Experiment 2). The majority of incorrect responses to multiple wh-questions given by SLI children were subject lists (41%), followed by object lists (20%). However, it was not the case that children who made use of exhaustive lists in single wh-questions 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 in their interpretation one wh-pronoun was not represented. The preference for subject lists in the SLI group is in contrast to two recent studies on typical acquisition (Heizmann, 2008; Oiry and Roper, 2009), which found many TD children (mostly at age 4) to respond to paired wh-questions with a list of objects. Oiry and Roper (2009) take their result for English to suggest that in comprehension children treat the first wh-word incorrectly as a scope-marker for the second, an option not available in English but in German (*Was glaubt Maria, wen sie besucht?* 'What does Maria think whom she'll visit?').[34] Consequently, children would answer a paired question such 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 homophonic with the question word (for example *was* 'what' vs. *(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.

Contrary to our prediction, paired wh-*und*-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 within one clause, but the interpretation of wh-words as pairs that is difficult.

[33] One of the reviewers questions what this result predicts for languages that do not have overt exhaustivity markers but plural marking. As *alles* does not force plural agreement and Dutch *allemaal* can combine with either singular or plural marking (see example (7b, c)), we suggest that plural marking in those languages does not work the same as in Dutch and German.

[34] This is comparable to partial movement answers for wh-questions such as *Who$_i$ did the boy ask t$_i$ what$_j$ to buy t$_j$?* where the medial wh-word is answered (de Villiers et al., 1990, 2011).

Evaluating SLI children's performance in light of the Missing Quantifier Hypothesis stated in (30), in Experiment 1 we provided evidence that many children with SLI did not recognize that the question domain has to be exhausted in single and multiple wh-questions, extending previous findings (Roeper, 2004; Schulz and Penner, 2002). Note that the children with SLI exhibited difficulties with general sentence comprehension and working memory (assessed in the SETK 3-5), but had no lexical deficit regarding the lexical items used in the wh-experiment and, crucially, had no difficulty interpreting singleton wh-questions (Pretests 1 and 2). Thus, their difficulty with exhaustive wh-questions seems to be semantic in nature (see also Schulz, 2010b). 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, see Roeper, 2007; Roeper et al., 2005). At the same time, pragmatic phenomena that require fixing underspecified meanings such as 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 provided an account for 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 with previous research into an acquisition path, the following tentative steps towards mastering exhaustivity can be formulated for typical development:

(33)
I. Constant interpretation of all single *wh*-questions
II. Overt *alles* is interpreted as exhaustive in single *wh*-questions
III. Single wh-questions are interpreted as exhaustive
IV. Paired wh-questions are interpreted as requiring exhaustive PL answers

These steps are tentative in several ways: It may be that overt *alles* is in fact mastered at the same time at which wh-expressions in single wh-questions are interpreted as a constant (33-I/II). It may be that there is a stage at which all single wh-questions are overgeneralized as exhaustive and no mention-some answers are allowed or that the ambiguity is recognized early on (33-III). Lastly, it is unclear when and how plural readings in single wh-questions emerge. Furthermore, we do not know if pairing can be a pragmatic response independent of a wh-expression. If, in the experiment above the question were *What's happening?* instead of *Who is eating what?*, would the children spontaneously provide paired list answers with 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 in island environments (Cheng and Demirdache, 2010; Krifka, 2001). Thus, further research is called for to substantiate and extend this developmental path. As a first step, in ongoing work, we explore whether the acquisition order of single before paired wh-questions can be established for younger typically developing 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, 2010a).

In conclusion, in this study we argued on both 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: Answers to multiple wh-questions are in the unmarked case obligatorily exhaustive in the adult grammar, and single pair and plural pair answers are not allowed. Instead we argued for a unified semantic account where exhaustivity is rooted in the question meaning, and exhaustivity in single and multiple wh-questions is treated within the same framework. This notion is captured by the idea that a feature is present on the wh-word, like the quantifier *every*, universally exhausting the question domain. The main findings from our two experiments were that exhaustivity appears in a systematic way, first in single and then in paired wh-questions, and that plural errors were absent in both SLI and TD children across different ages. These results we took to be compatible with the unified semantic account. Note that if single and paired wh-questions are unrelated, the null hypothesis is that they show no pattern in acquisition. 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 the properties exhaustivity and pairing are distinct because the pairing appears independently and later than exhaustivity. Regarding the implications for SLI, we suggest that the difficulties with exhaustivity we found in children with SLI are an indicator for a general problem with quantification, pointing to an underappreciated disorder related to semantics.

Acknowledgements

This research and preparation of the manuscript were supported by the National Science Foundation (BCS-0527509, SBE-0834032, Co-PI: Tom Roeper) and the Hessen State Grant for Centers of Excellence (LOEWE center IDeA, Co-PI: Petra Schulz). For helpful discussions of exhaustivity in wh-questions we thank Katharina Hartmann, Bart Hollebrandse, Cécile Meier, Barbara Pearson, Anja Müller, Uli Sauerland, Ede Zimmermann, and Malte Zimmermann. We thank the two anonymous reviewers for valuable comments and questions that helped clarify the ideas presented in this paper. We are grateful to Anja Müller for help with the data analysis and to Tim Hirschberg for help with preparation of the manuscript. Any remaining

errors are of course our own. Above all, we are indebted to Celia Jakubowicz for inviting us to present parts of this research at the EUCLIDES conference and for her inspiring support.

Appendix A

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

	Child	SETK subtest				
		VS (*T*-value)	MR (*T*-value)	PGN (*T*-value)	GW (raw value)	SG (*T*-value)
TD	DA	59	58	55	8	79
	MO	53	63	49	5	57
	GF	72	69	68	5	62
	NA	65	63	61	5	68
	LE	**39**[a]	43	46	6	67
	VA	65	72	68	5	63
	ML	49	**21**	49	5	59
	KO	**33**	63	41	6	66
	LE	53	55	61	6	63
	NI	49	66	55	4	62
	JO	60	56	61	4	54
	LU	59	72	61	5	74
	JA	53	66	43	6	71
	RA	59	55	49	4	64
	CA	53	58	49	5	68
	AR	53	63	43	6	63
	LA	59	79	55	7	68
	GE	49	44	**35**	4	54
	LO	46	49	55	6	53
	LA	59	44	49	7	46
TD mean		54.35	57.95	52.65	5.45	63.05
SLI	SA	**33**	44	**31**	5	47
	DE	**36**	49	41	4	**36**
	MA	**34**	41	**39**	6	**35**
	FR	**39**	55	**37**	6	49
	AL	43	43	**39**	4	**34**
	SB	**22**	**35**	**35**	5	**37**
	HE	49	**37**	**37**	5	**35**
	VI	**33**	44	**27**	3	**26**
	KE	**29**	41	**26**	4	**28**
	JA	**39**	44	**23**	4	43
	AI	**39**	48	**35**	4	**37**
	JB	**33**	**35**	**31**	3	**36**
	SO	46	49	**37**	3	**33**
	LP	**29**	**35**	**27**	3	**28**
	VI	49	43	**23**	**2**	**39**
	TH	**36**	53	43	4	**39**
	FY	**39**	46	**35**	3	**39**
	PA	**29**	44	**27**	3	40
	RI	**39**	44	**37**	4	**35**
	JU	53	**37**	**23**	5	58
SLI mean		37.45	43.35	32.65	4	37.7

[a] 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.

Child	SETK subtest		
	VS	MR	SG
MA	59	58	54
JO	59	60	54
LE	49	55	46
LU	59	60	50
AM	49	58	58

Appendix B (*Continued*)

Child	SETK subtest		
	VS	MR	SG
AN	43	51	55
MA	49	**39**[a]	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

[a] Bold numbers mark below average performance.

References

Beck, S., Rullmann, H., 1999. A flexible approach to exhaustivity in questions. Natural Language Semantics 7, 249–298.

Bošković, Z., 1998. Wh-phrases and wh-movement in Slavic. Position Paper presented at the Comparative Slavic Morphosyntax Workshop, Bloomington, Indiana. Online available, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.72.9721&rep=rep1&type=pdf.

Bošković, Z., 2003. On the interpretation of multiple questions. Linguistic Variation Yearbook 1, 1–15.

Cheng, L., Demirdache, H., 2010. Trapped at the edge: on long-distance pair-list readings. Lingua 120, 463–484.

Citko, B., 2008. Wh-questions with coordinated wh-pronouns. Abstract for the 29th TABU Dag, Groningen. Online available, http://www.let.rug.nl/~coltekin/tabu2008/abstracts.html#Citko.

Comorovski, I., 1996. Interrogative Phrases and the Syntax-Semantics Interface. Kluwer, Dordrecht.

Costa, J., 2004. Subject Positions and Interfaces: The Case of European Portuguese. Mouton de Gruyter, Berlin.

Dayal, V., 1996. Locality in Wh-quantification: Questions and Relative Clauses in Hindi. Kluwer, Dordrecht.

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

Dayal, V., 2005. Multiple wh questions. In: Everaert, M., van Riemsdijk, H. (Eds.), The Blackwell Companion to Syntax 3. Blackwell Publishers, Malden, pp. 275–327.

de Villiers, J.G., 1995. Empty categories and complex sentences: the case of wh-questions. In: Fletcher, P., MacWhinney, B. (Eds.), Handbook of Child Language. Blackwell Publishing, Oxford, pp. 508–540.

de Villiers, J.G., 2003. Defining SLI: a linguistic perspective. In: Levy, Y., Schaeffer, J. (Eds.), Language Competence across Populations. Toward a Definition of Specific Language Impairment. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 425–447.

de Villiers, J.G., Roeper, T., Vainikka, A., 1990. The acquisition of long-distance rules. In: Frazier, L., de Villiers, J.G. (Eds.), Language Processing and Language Acquisition. Kluwer, Dordrecht, pp. 257–297.

de Villiers, J.G., de Villiers, P., Roeper, T., 2011. Wh-questions: moving beyond the first phase. Lingua 121, 352–366.

De Vincenzi, M., et al., 1999. Parsing strategies in children comprehension of interrogative sentences. In: Bagnara, S. (Ed.), Proceedings of ECCS'99 (European Conference of Cognitive Science). Istituto di Psicologia del CNR, Rome, pp. 301–308.

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.

Ebbels, S., van der Lely, H.K.J., 2001. Meta-syntactic therapy using visual coding for children with severe persistent SLI. In: International Journal of Language and Communication Disorders 36 (supplement). pp. 345–350.

Friedmann, N., Novogrodsky, R., 2011. Which questions are most difficult to understand? The comprehension of wh questions in three subtypes of SLI. Lingua 121, 367–382.

Friedmann, N., Belletti, A., Rizzi, L., 2009. Relativized relatives: types of intervention in the acquisition of A-bar dependencies. Lingua 119, 67–88.

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

Grebenyova, L., 2006b. Multiple interrogatives in child language. In: Bamman, D., Magnistkaia, T., Zaller, C. (Eds.), BUCLD 30: Proceedings of the 30th annual Boston University Conference on Language Development. Cascadilla Press, Somerville, MA, pp. 225–236.

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

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

Grohmann, K.K., 2002. Multiple *wh*-fronting and the left periphery: German = Bulgarian + Italian. In: Mauck, S., Mittelstaedt, J. (Eds.), Georgetown University Working Papers in Theoretical Linguistics 2. pp. 83–115.

Grohmann, K.K., 2003. German is a multiple wh-fronting-language! In: Boeckx, C., Grohmann, K.K. (Eds.), Multiple Wh-Fronting. John Benjamins, Amsterdam, pp. 99–130.

Hagstrom, P.A., 1998. Decomposing Questions. Ph.D. dissertation, MIT.

Hagstrom, P.A., 2003. What Questions Mean. GLOT International 7/8, pp. 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: Aboh, E., Hartmann, K., Zimmermann, M. (Eds.), Focus Strategies in African languages: The Interaction of Focus and Grammar in Niger-Congo and Afro-Asiatic. Mouton de Gruyter, Berlin, pp. 241–263.

Heizmann, T., 2007. Children's acquisition of exhaustivity in clefts. In: Caunt-Nulton, H., Kulatilake, S., Woo, I.-h. (Eds.), BUCLD 31: Proceedings of the 31st annual Boston University Conference on Language Development. Cacadilla Press, Somerville, MA, pp. 298–309.

Heizmann, T., 2008. Exhaustivität in Spaltkonstruktionen, Fragen und Quantoren bei deutschen und englischen Kindern. Paper presented at Interdisziplinäre Tagung über Sprach-entwicklungsstörungen (ISES 5), University of Mainz. Online available, http://ises5.fh-fresenius.de/images/stories/nachberichte/ISES5_Heizmann_2008.pdf.

Hollebrandse, B., 2003. Distributive qualities of wh-words in Dutch child language. Paper presented at Taalkunde in Nederland dag (TIN-dag), Utrecht.

Jakubowicz, C., 2011. Measuring derivational complexity: new evidence from typically developing and SLI learners of L1 French. Lingua 121, 339–351.

Jakubowicz, C., Gutierrez, M.J., 2007. Elicited production and comprehension of root wh questions in French and Basque. Presentation at the COST Meeting Cross linguistically robust stage of children's linguistic performance, Berlin.

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: Fery, C., Sternefeld, W. (Eds.), Audiatur vox sapentia. A Festschrift for Arnim von Stechow. Akademie Verlag, Berlin, pp. 287–319.
Leonard, L.B., 1998. Children with Specific Language Impairment. MIT Press, Cambridge, MA.
Link, G., 1983. The logical analysis of plurals and mass terms: a lattice-theoretical approach. In: Bäuerle, R., Schwarze, C., von Stechow, A. (Eds.), Meaning, Use and Interpretation of language. Walter de Gruyter, Berlin, pp. 302–323.
Marinis, T., 2011. On the nature and cause of specific language impairment: a view from sentence processing and infant research. Lingua 121, 463–475.
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 University Press, Oxford.
Nelken, R., Shan, C.-C., 2004. A logic of interrogation should be internalized in a modal logic for knowledge. Paper presented at the Conference on Semantics and Linguistic Theory (SALT 14), Chicago. Online available, http://semanticsarchive.net/Archive/mEwOTFlO/nelken-shan-salt14.pdf.
Nishigauchi, T., 1999. Some preliminary thoughts on the acquisition of the syntax and semantics of wh-constructions. Theoretical and Applied Linguistics at Kobe Shoin 2, 35–48.
Oiry, M., Roper, T., 2009. How language acquisition reveals minimalist symmetry in the wh-system. In: Grohmann, K.K., Phoevos, P. (Eds.), Selected Papers from the Cyprus Syntaxfest. Cambridge Scholars Publishing, Cambridge, pp. 11–28.
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 (Nr. 77).
Penner, Z., 1999. Screeningverfahren zur Feststellung von Störungen in der Grammatikentwicklung. Edition SZH/CSPS, Luzern.
Penner, Z., Kölliker Funk, M., 1998. Therapie und Diagnose von Grammatikstörung: Ein Arbeitsbuch. Edition SZH/SPC, Luzern.
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 at the Graduiertenkolleg Satzarten. University of Frankfurt.
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: Tracy, R. (Ed.), Who Climbs the Grammar Tree. Niemeyer Verlag, Tübingen, pp. 465–492.
Roeper, T., 2004. Diagnosing language variations: underlying principles for syntactic assessment. Seminars in Speech and Language 25, 41–55.
Roeper, T., 2007. The Prism of Grammar: How Child Language Illuminates Humanism. MIT Press, Cambridge, MA.
Roeper, T., 2009. The minimalist microscope: how and where interface principles guide acquisition. In: Chandlee, J., et al. (Eds.), BUCLD 33: Proceedings of the 33rd annual Boston University Conference on Language Development. Cascadilla Press, Somerville, MA, pp. 24–48.
Roeper, T., de Villiers, J., 1991. The emergence of bound variable structures. In: Maxfield, T.L., Plunkett, B. (Eds.), Papers in the Acquisition of WH. Proceedings of the University of Massachusetts Roundtable (UMOP) GLSA Publications, Amherst, MA, pp. 225–266.
Roeper, T., de Villiers, J., 1993. The emergence of bound variables. In: Reuland, E., Abraham, W. (Eds.), Knowledge and Language (Volume I). From Orwell's Problem to Plato's Problem. Kluwer, Dordrecht, pp. 105–139.
Roeper, T., Pearson, B.Z., Strauss, U., 2005. Disordered language and complex syntax: quantifier spreading across dialect and impairing status. Paper presented at the X. International Congress for the Study of Child Language (IASCL), Berlin.
Roeper, T., et al., 2007. From singleton to exhaustive: the acquisition of wh-. In: Becker, M., McKenzie, A. (Eds.), Proceedings of SULA 3: Semantics of Under-Represented Languages in the Americas. GLSA Publications, Amherst, MA, pp. 87–102.
Schulz, P. 2010a. Who answered what to whom? On children's understanding of exhaustive questions. Paper presented at the final Conference of COST Action A33, London.
Schulz, P., 2010b. Some notes on semantics and SLI. In: Costa, J., et al. (Eds.), Language Acquisition and Development. Proceedings of GALA 2009. Cambridge Scholars Press, Newcastle upon Tyne.
Schulz, P., Penner, Z., 2002. The emergence of exhaustivity in wh-questions. Evidence from normal and impaired language acquisition in German. Handout presented at the Workshop on "Cooperations in Research on Specific Language Impairment", Groningen.
Seymour, H.N., Roeper, T., de Villiers, J., 2000. Dialect Sensitive Language Test (DSLT): Pilot Edition. The Psychological Corporation, San Antonio, Texas.
Siegmüller, J., Herzog, C., Herrmann, H., 2005. Syntaktische und lexikalische Aspekte 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.
Stromswold, K., 1995. The acquisition of subject and object wh-questions. Language Acquisition 4, 5–48.
van der Lely, H.K.J., Jones, M., Marshall, C.R., 2011. Who did Buzz see somebody? Grammaticality judgement of wh-questions in typically developing children and children with Grammatical-SLI. Lingua 121, 408–422.
van Rooij, R., 2004. Questions and relevance. In: Bernardi, R., Moortgat, M. (Eds.), Questions and answers. Proceedings of 2nd CoLogNET-ElsNET Symposium. pp. 96–107, Online available, http://www.let.uu.nl/~ctl/workshops/CES03/On_line_Proceedings/Papers/van_rooy.pdf.
von Stechow, A., Zimmermann, E., 1984. Term answers and contextual change. Linguistics 22, 3–40.
Wachowicz, K.A., 1974. Against the universality of a single wh-question movement. Foundations of Language 11, 155–166.
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 from an interdisciplinary longitudinal study. Paper presented at the COST Action A33 Meeting, Lisbon. Online available, http://cost.zas.gwz-berlin.de/cost/lisbon/weissenborn-hoehle.pdf.
Zimmermann, M., 2007. Quantifying question particles in German: syntactic effects on interpretation. In: Puig-Waldmueller, E. (Ed.), Proceedings of Sinn und Bedeutung (SuB 11), Barcelona, pp. 627–641.
Zimmermann, M., in preparation. Quantifying question particles and the non-exhaustiveness of wh-questions. MS, Potsdam University.