Lingua 121 (2011) 408–422

Contents lists available at ScienceDirect

Lingua

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

Who did Buzz see someone? Grammaticality judgement of wh-questions in typically developing children and children with Grammatical-SLI

Heather K.J. van der Lely *, Melanie Jones, Chloë R. Marshall [1]

Centre for Developmental Language Disorders and Cognitive Neuroscience, DLDCN.org, London, United Kingdom

ARTICLE INFO

Article history:
Received 31 May 2007
Received in revised form 22 September 2010
Accepted 27 September 2010
Available online 11 November 2010

Keywords:
Language acquisition
G-SLI
Grammaticality judgement
Wh-questions

ABSTRACT

This paper tests claims that children with Grammatical(G)-SLI are impaired in hierarchical structural dependencies at the clause level and in whatever underlies such dependencies with respect to movement, chain formation and feature checking; that is, their impairment lies in the syntactic computational system itself (the Computational Grammatical Complexity hypothesis proposed by van der Lely in previous work). We use a grammaticality judgement task to test whether G-SLI children's errors in wh-questions are due to the hypothesised impairment in syntactic dependencies at the clause level or lie in more general processes outside the syntactic system, such as working memory capacity. We compare the performance of 14 G-SLI children (aged 10–17 years) with that of 36 younger language-matched controls (aged 5–8 years). We presented matrix wh-subject and object questions balanced for wh-words (*who/what/which*) that were grammatical, ungrammatical, or semantically inappropriate. Ungrammatical questions contained wh-trace or T-to-C dependency violations that G-SLI children had previously produced in elicitation tasks. G-SLI children, like their language controls, correctly accepted grammatical questions, but rejected semantically inappropriate ones. However, they were significantly impaired in rejecting wh-trace and T-to-C dependency violations. The findings provide further support for the CGC hypothesis that G-SLI children have a core deficit in the computational system itself that affects syntactic dependencies at the clause level.

1. Introduction

1.1. Outline

Wh-questions have long played an important role in linguistic theory and language acquisition (Crain and Thornton, 1998; de Villiers, 1991; de Villiers and Roeper, 1995; Manzini, 1992; Manzini, 1995; Rizzi, 1990; Stromswold, 1995). Their theoretical interest stems partly from the syntactic representation underlying wh-questions, which typically involves wh-movement, and partly from the developmental changes that occur during their acquisition. More recently, such interest has been extended to children with Specific Language Impairment (SLI). Our first investigation of wh-questions in children with SLI revealed that they have significant impairments with producing syntactic dependencies (movement) involving both wh-trace and T-to-C dependencies (van der Lely, 1998; van der Lely and Battell, 2003). There has since been a plethora of

* Corresponding author at: Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, United States. Tel.: +44 020 7490 7042.
E-mail address: h.vanderlely@dldcn.org (Heather K.J. van der Lely).
URL: http://www.dldcn.org
[1] Current address: Department of Language and Communication Science, City University London, United Kingdom.

0024-3841
doi:10.1016/j.lingua.2010.10.007

investigations of wh-questions in SLI in languages as diverse as English, Hebrew, French, Swedish, Greek, and Chinese (Deevy and Leonard, 2004; Friedmann and Novogrodsky, 2007; Hamann, 2006; Hansson and Nettelbladt, 2006; Stavrakaki, 2006; Wong et al., 2004). Where the language investigated incurs wh-trace dependencies, wh-questions are uniformly found to be significantly impaired in children with SLI. Such studies provide a new window on the acquisition of questions, as well as furthering our understanding of the underlying nature of SLI.

This paper forms part of a body of work in our lab investigating wh-questions using different methodologies to explore different input and output representations and processes in typically developing children and a subgroup of SLI children with Grammatical(G)-SLI (see also Fonteneau and van der Lely, 2008; Marinis and van der Lely, 2007; van der Lely and Battell, 2003). Here we use a grammaticality judgement task to tap input processes and representations of linguistic structures. This methodology has been shown to be particularly insightful when investigating subjects with acquired language disorders (aphasia) (Tyler, 1992), because it allows the researcher to distinguish impairments in representations/stored syntactic knowledge from impairments that occur later in the processing chain to full comprehension or production. Such processes include working memory, processing speed or capacity, processing at the interface, and knowledge outside the syntactic system (pragmatics/world knowledge). Thus, the aim of this study is to further distinguish whether the SLI children's impairment in wh-questions lies in the syntactic computational system itself (Friedmann and Novogrodsky, 2007; van der Lely, 2005; van der Lely and Battell, 2003), or outside the language faculty (Deevy and Leonard, 2004; Jakubowicz and Strik, 2008; Stavrakaki, 2006; Wong et al., 2004). Before discussing some of the different theories proposed for SLI, we discuss previous research that has investigated wh-questions in children with SLI.

1.2. Wh-questions

We focus on the simple matrix subject and object questions that have been the topic of much debate in syntactic theory (Manzini, 1992; Manzini, 1995; Rizzi, 1990; Stromswold, 1995). It is generally agreed that in English, object wh-question formation involves two forms of syntactic dependencies as defined by syntactic movement. The first is movement of the wh-operator to the specifier (spec) position of the complementizer phrase (CP), which leaves a trace behind in the internal verb argument position, as in (1) (Rizzi, 1990). This prevents, in adult grammar, the empty internal verb argument position being filled by a determiner phrase (DP), as shown in (2). Second, object questions necessitate T-to-C dependency (or "*do*-support") of *do* bearing the question-feature to the head of CP, (1). *Do*-support determines appropriate tense and question-feature marking in object questions (hereafter T-to-C dependency).

(1)[$_{CP}$ Who$_i$[$_{C}$ did$_j$][$_{TP}$Homer [$_{T}$t$_j$]][$_{VP}$ [$_{V}$ find] t_i[$_{PP}$ at the farm]]]

(2) * [$_{CP}$ Who$_i$[$_{C}$ did$_j$][$_{TP}$ Homer [$_{T}$ t$_j$]][$_{VP}$ [$_{V}$ find] Bart$_i$ [$_{PP}$ at the farm]]]

For subject questions we assume the analysis of Rizzi (1996), whereby the wh-word moves from an original position within the inflectional phrase (IP) to the CP as shown in (3) (but see Pesetsky, 1987 for contrasting analysis). In contrast to object questions, subject questions do not incur *do*-support, and therefore no T-to-C dependency occurs. Tense, however, is typically marked on the matrix verb following covert V-to-T movement.[2]

(3)[$_{CP}$ Who$_i$[$_{C'}$ [$_{TP}$ t$_i$ [$_{T'}$ [$_{VP}$ [$_{V'}$ found Homer[$_{PP}$ at the farm]]]]]]]

Despite their syntactic complexity, typically developing children acquire questions early in language development, acquire object questions at the same time as subject questions (around 3 years or earlier), and even show early acquisition of more complex long-distance wh-questions (Stromswold, 1995; Thornton, 1990, 1995). This competence is robust across languages despite variations in vocabulary and features to be learnt (Hamann, 2006; Jakubowicz and Strik, 2008; Stavrakaki, 2006; Weissenborn et al., 1995). This is not so, however, for children with SLI.

There is considerable theoretical discussion concerning what underlies and creates the syntactic dependencies such as those involved in wh-questions, e.g., feature specification, feature checking, movement, chains, merge and agree. What is evident is that the dependencies come in pairs and their interpretation is "blind" to semantics. However, for the purposes of this paper we are glossing over these issues, as what is important here is the phenomenon itself in relation to clause structure.

[2] Please note that for the purpose of this paper we use the notation T to C without any commitment to theoretical discussions of Split IP, etc.

1.3. SLI and wh-questions

SLI is an impairment in acquiring language despite otherwise normal intelligence, hearing and an adequate learning environment (Leonard, 1998). The disorder heterogeneously affects comprehension and production in components of language such as syntax, morphology, phonology, vocabulary, and in some children, pragmatics. The deficit can persist into adulthood (Clegg et al., 2005; Conti-Ramsden and Durkin, 2007; van der Lely, 2005). It has a strong genetic component (Fisher et al., 2003), and two genes (FOXP2, CNTNAP2) and several loci have been linked to different forms of SLI (Bishop et al., 2006; Marcus and Fisher, 2003; Vernes et al., 2008). In this study we focus on children with G-SLI (cf. "Syntactic-SLI", Friedmann and Novogrodsky, 2007). G-SLI is a sub-group of SLI characterised by a persistent and primary deficit in grammar (van der Lely, 2005; van der Lely et al., 1998, 2004). Preliminary evidence reveals familial clustering of language impairment consistent with an autosomal dominant inheritance (van der Lely and Stollwerck, 1996). However, the nature of the language impairment in family members varies, suggesting a more complex inheritance. G-SLI manifests itself as impairments in syntax, morphology and phonology (Gallon et al., 2007; Marshall and van der Lely, 2006, 2007; van der Lely, 1998, 2005; van der Lely and Marshall, 2010). More specifically, we have argued that G-SLI affects hierarchical structural complexity in syntax, morphology and phonology (van der Lely, 2005). In this paper, we explore the nature of the syntactic deficit.

The syntactic characterisation of G-SLI includes errors in tense marking (*Joe go home*), as has been elegantly shown by Rice and colleagues for other groups of children with SLI (Rice and Wexler, 1996a). However, the deficit extends to the assignment of theta roles to noun phrases in active and passive sentences (*Joe was hit by Bill*) (van der Lely, 1996a), and to pronouns and reflexives (*Mowgli says Baloo is tickling him/himself*) when contextual or pragmatic cues cannot determine assignment (van der Lely and Stollwerck, 1997). Of particular relevance to this paper is the impairment in wh-questions (van der Lely, 1998; van der Lely and Battell, 2003).

Our initial investigation of wh-questions in G-SLI explored subject and object matrix questions, such as those in (1) and (3), using an elicitation task based on a "whodunnit?" game (van der Lely and Battell, 2003). The study was designed to investigate the hypothesis that G-SLI children have a core deficit in movement (head-to-head movement and operator movement). This study revealed that teenaged G-SLI participants are impaired on both subject and object questions. As hypothesised, due to the additional movement operations in object questions in comparison to subject questions, they were more impaired on object than subject questions. Furthermore the pattern of errors was in just those areas that would be predicted if SLI children are impaired in syntactic dependencies. That is, they showed both types of movement errors: wh-movement errors, such as "gap-filling" (4a) and moving only the wh-operator part of the wh-phrase (4b), and T-to-C dependency errors, such as omission of do-support (5a) and double tense marking (5b):

(4) a. *Who did Mr Green saw somebody?
b. *Which did Mrs Peacock like jewellery?
(5) a. *What cat Mrs White stroked?
b. *What did she spotted in the library?

Such errors are found not only in English-speaking children with SLI, but also in French, Greek, German, Italian, and Hebrew for wh-questions and relative clause constructions (Adani et al., submitted for publication; Friedmann and Novogrodsky, 2004; Hamann, 2006; Hamann et al., 1998; Jakubowicz et al., 1998; Stavrakaki, 2001). Furthermore, additional error types, such as wh-in-situ preference in French and case errors in Greek, are attested cross-linguistically. Interestingly, such errors can also be accounted for by an impairment in movement (Hamann, 2006; Stavrakaki, 2006). Papers in this special issue eloquently provide details of the most recent studies in this area, and the reader is directed to these papers (de Villiers et al., 2011; Friedmann and Novogrodsky, 2011; Jakubowicz, 2011; Schulz and Roeper, 2011.

Following up on our initial elicitation data, van der Lely and colleagues used an on-line cross-modal priming study to tease apart the representations/mechanisms that were implemented during the on-line processing of questions (Marinis and van der Lely, 2007). Specifically, we were interested in the underlying representations/mechanisms involved in the syntactic dependency between a wh-word and the gap in indirect object questions, that is, the "filler-gap" dependency, as shown in (6):

(6) a. Bart gave the long carrot to the *rabbit*

b. [$_{CP}$ *Who$_i$* did Bart give the long carrot to t_i at the farm]

As children listened to sentences such as the example in (6), a picture of a *rabbit* (target noun) or *ladder* (control noun) appeared in one of three different positions: at either the offset of the verb, *give* (where we anticipated lexical-priming of associated arguments), the off-set of the adjective, *long* (a control position, where no priming was anticipated), or the position of the trace (gap) where priming was expected. The child's task was to push a button corresponding to whether the picture was animate or inanimate. Therefore if he/she reactivated the noun in the trace position, a faster reaction time (priming) would be recorded relative to the control noun. For the language-matched and the memory- and age-matched control groups we found significant priming at the trace position, indicating reactivation at the trace. In contrast, for the children with G-SLI priming was revealed at the offset of the verb, but not at the trace. The findings indicated that the G-SLI children were interpreting wh-questions via lexical-thematic information, and that they failed to reliably establish a syntactic filler-gap dependency (Marinis and van der Lely, 2007). A further study used Event Related Potentials in order to measure brain responses on a millisecond by millisecond basis to establishing structural dependent relations between the animacy property of the wh-word (*who* vs. *what*) and the animacy property of the noun which filled the first possible gap after the verb (e.g., *clown* vs. *ball*) (Fonteneau and van der Lely, 2008). Previous piloting with over 700 typically developing children showed that it was unexpected for the animacy properties of a wh-word and filled gap to match (either both animate or both inanimate). We were interested in, not so much that children would notice these properties, but which neural mechanisms would use this information to facilitate processing. We found that typically developing age and language matched children showed a very fast (around 150–300 ms) neural correlate, known as the Early Left Anterior Negativity (ELAN) that is associated with automatic syntactic structural knowledge. The G-SLI children did not show this response at all, but a later response around 400 ms that is typically associated with semantic or pragmatic processing (Fonteneau and van der Lely, 2008). Thus, it appeared that the G-SLI participants were compensating for their syntactic deficit, semantically. Furthermore, the individual ERP data revealed a remarkably consistent pattern across the children in the G-SLI group. Interestingly, all children showed a normal later response known as the P600 that is associated with re-analysis and is not found to be language specific.

However, what we do not know from these studies is whether, if we reduced or changed the general processing load (not to be confused with any processing load that is specific to the syntactic representations themselves), G-SLI children would exhibit evidence of normal representations of syntactic dependencies. In other words, is the deficit in the implementation of the syntactic dependency operation itself or is it outside syntax proper, in limitations of general working memory capacity or processing capacity (Gathercole, 2006; Jakubowicz and Strik, 2008; Wong et al., 2004)?

We contribute to this issue by investigating whether G-SLI children and younger language-matched control children recognize an ungrammatical question containing the very errors that they produce in elicitation tasks (either wh-trace or T-to-C dependency errors), and comparing their performance with recognition of questions containing semantic errors (verb subcategorisation or wh-word animacy (i.e. *who* for *what*) errors). We aim to tap syntactic input processes and stored representations, rather than the additional more general processes required for full comprehension or production. We first discuss some current theories of SLI and the predictions they would make for this study.

1.4. Some current theories of SLI

The discussion below is limited to hypotheses that can explicitly account for the deficits in syntactic dependency operations at the clause level that are evident in wh-questions in SLI children. Thus, hypotheses that aim to provide a clinical marker of SLI, and by definition account for only part of SLI grammar, will not be discussed here, e.g., the Extended Optional Infinitive account (Rice and Wexler, 1996b) or the Agreement Deficit hypothesis (Clahsen et al., 1997).

To account for the broad range of syntactic deficits in G-SLI, van der Lely and colleagues have developed the Computational Grammatical Complexity (CGC) hypothesis, which is a development and extension of the Representational Deficit for Dependent Relations (RDDR) hypothesis (Marshall and van der Lely, 2007; van der Lely, 1998, 2005). The CGC hypothesis provides a framework for characterizing the deficits in syntax, morphology and phonology that are typically found in many forms of SLI, not only G-SLI. Our work reveals that although there are many ways of increasing grammatical complexity, many school-aged SLI children lack the computations to consistently form hierarchical, structurally complex forms in one or more of the components of grammar that normally develop between 3 and 6;6 years (Gallon et al., 2007; Marshall and van der Lely, 2006; van der Lely, 1998; van der Lely and Battell, 2003; van der Lely and Christian, 2000; van der Lely and Marshall, 2010; van der Lely and Ullman, 2001). This working hypothesis emphasizes the notion that impairments in syntax, morphology and phonology are functionally autonomous, but cumulative in their effects. This view predicts that SLI should arise from a number of deficits—some specific, some not. Moreover, deficits in the different components of grammar might co-exist or dissociate.

Here we will provide further details of the syntactic aspect of the CGC hypothesis, which is a development of the original RDDR hypothesis (van der Lely, 1998; van der Lely and Stollwerck, 1997). The CGC hypothesis claims that the underlying syntactic deficit is in the syntactic computational system (representations and/or mechanisms).[3] Specifically the deficit is in syntactic dependencies (van der Lely, 1998, 2005; van der Lely and Marshall, 2010). Whereas the syntactic dependencies within the nominal phrase are thought to be normal, those in the clause are not. In other words, the implementation of syntactic dependencies at the clause level is impaired. This deficit prevents SLI children building (and therefore, logically too, also parsing) syntactically complex hierarchical structures involving clausal syntactic dependency operations. Thereby the probability of errors increases as a function of increasing the number of clausal dependency operations, e.g., as in passives, or embedded structures. Central to this hypothesis is that the deficit in syntactic dependencies is in the computational syntactic system itself, rather than in processes outside the syntactic system. It is an empirical issue that goes beyond the scope of this paper to disentangle the precise location of the deficit, and indeed the relations between the underlying properties that are

[3] To avoid ambiguity, unless otherwise specified when we refer to "mechanism" we are referring to the neural implementation of the syntactic dependency. We consider that this is part of the domain-specific computational syntactic system (van der Lely, 2005). We are not making any commitment to "modularity". Further, we are not referring to any general mechanism(s) outside the syntactic system.

needed to be in place for accurate clausal syntactic dependencies to occur; e.g., grammatical feature specification and feature requirements of the constituents involved, and the correct parsing of these constituents. In some current linguistic terms the operation ''Agree'' forms the dependency, with the ensuing movement occurring as a ''side-effect''. For the purpose of this paper we hypothesise that clausal syntactic dependencies are impaired, without committing to the precise location within the set of necessary conditions and operations. However, the afflicted individual can use alternative resources outside the syntactic system, such as the semantic system, to compensate wherever possible. For example semantically interpretable features could facilitate relations between constituents through probabilistic association mechanisms, thereby invoking different mechanisms and representations that those underlying syntactic dependencies, but nonetheless facilitating sentence processing. The use of compensatory mechanisms is supported by findings from our cross-modal priming (Marinis and van der Lely, 2007) and Event Related Potential (Fonteneau and van der Lely, 2008) studies. In the latter study, electrophysiological brain responses revealed a selective impairment to neural circuitry that is specific to syntactic structural dependencies, and that G-SLI participants partially compensated by using neural circuitry associated with semantic processing.

Recently, Friedmann and colleagues investigated the deficit in movement more precisely by studying Hebrew speaking children with Syntactic(S)-SLI (Friedmann and Novogrodsky, 2007). Using an ingenious reading paradigm of homographs in Hebrew, they were able to distinguish whether the movement deficit was in deriving the syntactic structure and traces, or in the assignment of thematic roles from those traces to the moved elements. The results suggested that the deficit was in the latter and that this led to impairment in the comprehension of movement-derived sentences. Note here too the deficit is defined as being in the computational system itself.

Whereas van der Lely and Friedmann argue for economy of structure caused through impaired syntactic dependencies negotiated through feature requirements, agree and movement, Rizzi argues that in early grammar children simply do not build a full structure with CP—the Truncation Hypothesis (Rizzi, 2000). Thus, children use the minimal structure that can accommodate the overt material. Rizzi's hypothesis is not explicitly a hypothesis about impaired grammar—it aims to characterise typical development at an early stage, before full syntactic structure is available. However, if children with SLI have not yet acquired a particular structure, then their grammar will resemble that of younger, typically developing children; therefore, parallels might be drawn between constraints that prevent both grammars being adult-like.

A related hypothesis is the Derivational Complexity Hypothesis (DCH), developed by Jakubowicz and colleagues. The DCH differs from the CGC hypothesis in interesting ways (Jakubowicz and Strik, 2008). Jakubowicz argues for a deficit in syntactic complexity affecting both merge and movement. She proposes that as movement is more costly, this will be impaired to a greater extent than merge, and the more movement operations that are required the greater the impairment. In this respect the CGC and the DCH are very similar in the predictions that they make. The DCH, however, differs from the CGC hypothesis in that it argues that the deficit lies in limitations of working memory or processing resources, which impact on syntactic derivations (cf. Wong et al., 2004). The issue of working memory and processing resources raised by Jakubowicz are highly pertinent. However, a distinction needs to be drawn between domain-general working memory and processing resources and working memory and processing that may be intrinsic and specific to the syntactic dependencies we are investigating. This study is not designed to explore these issues and further discussion goes beyond the scope of this paper, so we leave this to future research.

The hypotheses discussed so far lead to a number of predictions. First, they all predict that grammaticality judgements of G-SLI children will be impaired for questions containing syntactic dependency violations at the clause level. However, the CGC hypothesis predicts that if object and subject questions contain the same number of dependency violations, then they should be impaired to the same extent. An alternative prediction is that object questions are more difficult than subject questions due to the greater distance of the dependency between the wh-word and the trace, which requires the wh-word to be held in working memory working memory for longer before it can be interpreted. Hence the processing demands of object questions are predicted to be greater. If this is limited in G-SLI, as proposed by the DCH, then object questions should be more impaired than subject questions.[4] Note that our interpretation of the DCH is that the working memory and processing discussed by Jakubowicz is of a general nature rather than specific to syntactic representations. Further, the DCH might predict that if G-SLI children, like adults with aphasia, have unimpaired representations, but are impaired in general resources needed to build those representations, then the grammaticality judgement task could reveal normal or near normal judgements. In contrast, the CGC hypothesis claims that the deficit is in representations/mechanisms pertaining to syntactic dependencies in the clause, so a deficit should be revealed regardless of the method of testing.

2. Method

2.1. Participants

Four groups participated in the study: children with G-SLI and three groups of younger control children matched on their grammatical or vocabulary abilities.

[4] An anonymous reviewer suggests that in the phase-based approach (Chomsky, 1999), working memory should not make a difference for subject and object questions since both wh-constituents will be in the same search space for C_Q (on the edge of the vP-phase). The relationship between working memory and linguistic knowledge is a complex one. Subject questions may be facilitated by cognitive strategies which are no help for object questions, as our ERP data show (Fonteneau and van der Lely, 2008).

Table 1
Summary of participants' details for the G-SLI and control groups.

	G-SLI (*N* = 14)		LA1 (*N* = 12)		LA2 (*N* = 12)		LA3 (*N* = 12)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Chronol. age	13;2	(2.2)	6;0	(0.4)	7;0	(0.4)	8;1	(0.4)
Age range	10;5–17;1		5;7–6;5		6;6–7;6		7;7–8;6	
TROG[a]	14.36	(1.60)	14.67	(2.27)	15.83	(1.90)	17.50	(14.36)
z-Score	−1.80	(0.67)	0.64	(0.92)	0.54	(0.75)	0.98	(1.20)
Equivalent age	7;6	(1.5)	7;9	(1.9)	8;10	(1.7)	10;0	(1.6)
GC-ITPA[b]	21.86	(3.88)	22.58	(4.70)	27.00	(2.73)	28.08	(2.97)
z-Score	n/a		1.58	(1.16)	1.76	(0.56)	0.94	(0.83)
Equivalent age	7;3	(1.0)	7;7	(1.4)	9;0	(1.0)	9;3	(0.11)
BPVS[c]	78.21	(16.81)	61.08	(6.35)	76.33	(15.18)	84.67	(14.33)
z-Score	−2.00	(0.56)	0.43	(0.44)	0.88	(0.95)	0.74	(0.98)
Equivalent age	8;7	(2.0)	6;7	(0.8)	8;4	(1.9)	9;2	(1.7)

[a] Test of Reception of Grammar.
[b] Illinois Test of Psycholinguistic Abilities
[c] British Picture Vocabulary Scales.

2.1.1. G-SLI participants

Fourteen participants with G-SLI aged between 10;5 and 17;1 (mean 13;2) were recruited from two specialist language schools in the UK. All children were classified as having G-SLI (van der Lely, 1996a), that is they had persistent problems with grammatical aspects of language but their non-verbal IQ was 85 or above and their articulatory motor abilities, speech intelligibility and social–emotional behaviour were age-appropriate.

We used three standardised tests to select children with SLI, followed by more specific tests tapping the core aspects of morpho-syntax that are characteristically impaired in G-SLI. The standardised tests were the Test of Reception Of Grammar (TROG) (Bishop, 1989), that measures general grammatical comprehension, the Illinois Test of Psycholinguistic Abilities (GC-ITPA) (Kirk et al., 1968), that measures expressive morphology, and the British Picture Vocabulary Scales (BPVS) (Dunn et al., 1982), a test of single word vocabulary comprehension. On the TROG, the G-SLI children had a mean *z*-score of −1.80 with an equivalent age of 7;6 (1.5 SD), on GC-ITPA they had a mean equivalent age of 7;3 (1.0 SD) and on the BPVS they had a mean *z*-score of −2.00 with an equivalent age of 8;7 (2.0 SD). Table 1 provides a summary of the participants' scores from the standardised tests.

In addition, the G-SLI participants were selected on the basis of their scores on two non-standardised tests that tap grammatical phenomena that characterise children with G-SLI, the Verb Agreement and Tense Test (van der Lely, 1999), tapping the expression of tense and agreement, and the Test of Active and Passive Sentences (van der Lely, 1996b), tapping the comprehension of reversible active and passive sentences. Each participant included in this study made 20% or more errors in each of these two tests. This group of G-SLI participants also participated in a number of other studies, including investigation of the production of wh-questions, negation, phonological representations and auditory processing (Davies, 2002; Gallon et al., 2007; van der Lely, 2004; van der Lely and Battell, 2003).

2.1.2. Language ability control groups

Three groups of typically developing children each consisting of 12 children provided language ability (LA) control groups. These were matched to the children with G-SLI on the basis of the TROG, GC-ITPA and BPVS. The children were randomly selected from a state school in England, and they fell within the normal range of abilities as assessed by the three standardised tests. The LA groups' scores from the standardised tests are provided in Table 1. The youngest LA1 group (mean age 6;0, range 5;7–6;5) did not differ from the G-SLI group on the tests tapping general grammatical abilities (TROG) and morphology (GC-ITPA). However, they scored significantly lower than the G-SLI group on the test of vocabulary comprehension (BPVS) [$t(17.12) = 3.53$, $p < .005$]. The LA2 (mean age 7;0, range 6;6–7;6) and the LA3 (mean age 8;1, range 7;7–8;6) groups did not differ from the G-SLI children on the BPVS. In contrast, they scored significantly higher than the G-SLI group on the TROG [LA2: $t(24) = 2.15$, $p < .05$; LA3: TROG $t(24) = 4.37$, $p < .001$], and the GC-ITPA [LA2: $t(24) = 3.84$, $p < .005$; LA3: $t(24) = 4.53$, $p < .001$].

2.2. Design and materials

The task required children to decide whether an auditorily presented wh-question sounded correct ("good") or incorrect ("bad"). We used two types of grammatical and ungrammatical wh-questions, subject (*Who kicked Cookie Monster?*) and object (*Who did Cookie Monster kick?*). We compared these two question types for *who*, *what* and *which* wh-words in a 2 (Qn-type: Subject, Object) × 3 (Wh-word: *who*, *what*, *which*) × 4 (Group: SLI, LA1, LA2, LA3) design. Ungrammatical sentences were further subdivided into questions with a +wh,−T error involving lack of *do*-support, lack of tense marking or double tense marking, and questions with a −wh,+T error involving gap-filling, gap-filling involving *which one* and lack of pied-piping. Table 2 provides examples for each condition for the grammatical and Table 3 for the ungrammatical

Table 2
Example stimuli for grammatical wh-questions.

Question type	Wh-word	Example
Subject	Who	*Who posted the letter?*
	What	*What lifted Charlie Brown?*
	Which	*Which giraffe liked the zebra?*
Object	Who	*Who did Mowgli hug?*
	What	*What did the tiger chase?*
	Which	*Which baby did the nurse wash?*

Table 3
Example stimuli for ungrammatical wh-questions.

Question type		Wh-word	Condition	Example
+wh,−T	Subject	Who	No Tense Marking	*Who kiss Miss Piggy*
		What		*What follow the rabbit?*
		Which		*Which boy listen at the door?*
		Who	Double Tense Marking	*Who did kicked Cookie Monster?*
		What		*What did barked at the postman?*
		Which		*Which Spice Girl did danced at the nightclub?*
	Object	Who	No Tense Marking	*Who Homer Simpson beg?*
		What		*What Eeyore bump?*
		Which		*Which cup Peter Pan drop?*
		Who	No *do-support*	*Who Tinky Winky tickled?*
		What		*What Postman Pat pushed?*
		Which		*Which door the policeman locked?*
		Who	Double Tense Marking	*Who did Barney Rubble watched?*
		What		*What did Superman carried?*
		Which		*Which ball did David Beckham kicked?*
−wh,+T	Subject	Who	Gap-filling	*Who someone watched Fred Flintstone?*
		What		*What something pushed the lion?*
		Which	Gap-filling + which one	*Which milkman one passed the house?*
				Which one Power Ranger attacked the monster?
	Object	Who	Gap-filling	*Who did Tweety Pie touch someone?*
		What		*What did Popeye move something?*
		Which	Gap-filling + which one	*Which Mutant Turtle did Batman help one?*
				Which one did the teacher wipe the blackboard?
			No pied-piping	*Which did Scooby Doo follow the ghost?*

wh-questions. To test whether children are able to identify semantic violations we used a set of semantically inappropriate wh-questions violating verb subcategorisation properties, or wh-word properties, e.g., *Which telephone did the sandwich rush? Who table did Baloo Bear frighten?* In total there were 176 wh-questions, with 74 (42%) subject questions and 102 (58%) object questions. This slight bias to object questions was made as there were more types of possible movement errors in object questions.

We used a wide range of characters and verbs to keep the subjects interested in the task, and no character or verb appeared in more than four questions. All verbs were high frequency regular verbs. The stimuli were randomly distributed across three game sessions, but presentation order was consistent across subjects.

2.3. Procedure

We presented the stimuli in three game sessions to maximise participants' interest and attention. In each of the first two games participants were introduced to three language-learning characters (three Aliens in Game 1, and three Foreign Spice Girls in Game 2). Participants were told that the characters were not sure how to ask questions and needed their help. Each character asked a question in turn and participants had to decide whether they thought the question sounded good or not. A slight emphasis was placed on the verb, to avoid any emphasis inadvertently being placed on the ungrammatical words. If children thought the question was correct they gave that character a star, and if they thought it was incorrect they gave the character a spider. At the end of the game the character with the most stars was the winner. The third game was a competition between the two winners of Games 1 and 2 (Alien vs. Spice Girl). We presented four practice questions at the beginning of testing to ensure that participants understood the task. The instructions used in each of the three games are given in Appendix A.

3. Results

Prior to the analyses we screened the data for outliers. The score of one LA1 participant was more than 2 standard deviations below the group mean, and was therefore excluded from the analyses. The overall mean correct response of the

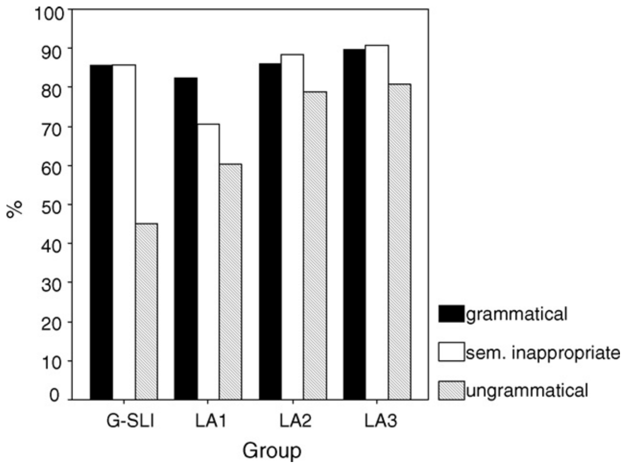

Fig. 1. Mean percentage of accurate responses in the three sentence types.

remaining participants was 76.6% (SD = 21.1). The percentage of correct responses for each of the groups in the grammatical, inappropriate and ungrammatical conditions is shown in Fig. 1.

3.1. *Analysis according to sentence type*

An initial 3 × 4 ANOVA with the factors *Sentence Type* (grammatical, ungrammatical, semantically inappropriate) and *Group* (G-SLI, LAl, LA2, LA3) was used to investigate whether G-SLI and TD children are able to judge equally well the grammaticality of the three sentence types. This analysis revealed a significant main effect of *Sentence Type* [$F(2, 90) = 29.532, p < .001$], a significant main effect of *Group* [$F(3, 45) = 5.170, p < .01$] and a significant interaction between *Sentence Type* and *Group* [$F(6, 90) = 5.813, p < .001$]. We followed up this interaction through one-way ANOVAs for each one of the three sentence types. For the grammatical and the semantically inappropriate sentences there was no significant difference between the groups [grammatical: $F(3, 45) = 1.059, p = .376$; semantically inappropriate: $F(3, 45) = 2.168, p = .105$]. In contrast, we found a significant difference between the groups in the ungrammatical sentences [$F(3, 45) = 11.108, p < .001$]. A Tukey HSD post hoc test revealed significant differences between G-SLI children and the two vocabulary control groups, LA2 [$p < .001$], and LA3 [$p < .001$], but no reliable difference between the children with G-SLI and the grammar controls, LA1 [$p = .178$]. There was also a significant difference between the younger (LA1) and the older (LA3) control groups [$p < .05$], but the difference between the LA1 and LA2 groups did not reach significance [$p = .082$].

3.2. *Analysis according to error type*

To investigate further the differences between the groups in the ungrammatical sentences and to identify the cause of errors, we analysed the ungrammatical sentences according to *Error Type* (+wh,−T, −wh,+T), *Sentence Type* (subject, object) and *Wh-Word* (*who*, *what*, *which*). The percentage of correct responses (and standard deviation) for each of the groups in the ungrammatical sentences are shown in Tables 4 and 5.

Table 4
Mean percentage (and standard deviations) of correct rejections of the ungrammatical questions.

	Question type	Wh-word	SLI	LA1	LA2	LA3	Total
+wh,−T	Subject	Who	48.2 (22.9)	47.7 (33.5)	80.2 (19.6)	84.4 (14.2)	65.1 (22.6)
		What	42.0 (18.7)	53.4 (30.2)	80.2 (21.0)	91.7 (13.4)	66.8 (20.8)
		Which	48.2 (20.1)	60.2 (30.0)	83.3 (25.7)	83.3 (25.7)	68.8 (25.4)
	Object	Who	28.6 (13.1)	54.5 (25.0)	76.2 (24.6)	81.0 (27.1)	60.1 (22.5)
		What	35.2 (16.4)	53.2 (25.0)	84.5 (21.7)	84.5 (27.2)	64.4 (22.6)
		Which	30.1 (16.6)	56.5 (28.7)	78.6 (24.4)	83.9 (30.0)	62.3 (24.9)
−wh, +T	Subject	Who	43.8 (23.4)	61.4 (34.7)	88.5 (17.2)	85.4 (31.5)	69.8 (26.7)
		What	52.4 (31.3)	71.2 (32.6)	76.4 (27.9)	69.4 (39.5)	67.4 (32.8)
		Which	60.7 (17.9)	73.4 (23.3)	91.1 (13.7)	86.9 (23.2)	78.0 (19.5)
	Object	Who	51.2 (29.6)	62.1 (35.0)	86.1 (15.6)	84.7 (35.9)	71.0 (29.0)
		What	54.8 (31.6)	68.2 (31.1)	94.4 (14.8)	90.3 (19.4)	76.9 (24.2)
		Which	50.8 (17.4)	73.7 (24.6)	86.6 (22.3)	86.1 (31.2)	74.3 (23.9)

Table 5
Mean percentage (and standard deviations) of correct rejections of the ungrammatical questions.

	Error type	SLI	LA1	LA2	LA3	Total
+wh,−T	No tense	39.3 (16.2)	63.6 (20.5)	76.7 (18.8)	78.5 (24.7)	63.4 (25.4)
	Double tense	38.9 (12.6)	55.0 (29.8)	85.0 (21.9)	87.7 (21.0)	65.8 (29.8)
	No do-support	30.6 (19.2)	50.0 (28.1)	73.4 (30.2)	81.5 (29.2)	57.9 (33.2)
−wh,+T	Gap-filling	55.4 (11.8)	68.9 (29.1)	86.8 (16.1)	83.1 (26.8)	72.9 (24.6)
	Which-one	45.8 (20.9)	68.2 (26.8)	88.9 (16.0)	87.5 (29.2)	71.6 (29.3)
	Pied-piping	51.2 (20.1)	77.3 (22.7)	86.1 (25.5)	87.5 (29.4)	74.5 (28.3)

A 2 × 2 × 3 × 4 ANOVA with the factors *Error Type* (+wh,−T, −wh,+T), *Question Type* (subject, object), *Wh-Word* (*who*, *what*, *which*), and *Group* (SLI, LA1, LA2, LA3) revealed a significant main effect of *Error type* [$F(1, 45) = 16.827$, $p < .001$], *Wh-Word* [$F(2, 90) = 3.213$, $p < .05$] and *Group* [$F(3, 45) = 12.549$, $p < .001$]. There were also significant interactions between *Error Type* × *Group* [$F(3, 45) = 3.024$, $p < .05$], *Error Type* × *Question Type* [$F(1, 45) = 4.192$, $p < .05$] and *Question Type* × *Wh-Word* [$F(2, 90) = 4.865$, $p = .01$]. The interaction between *Error Type* × *Question Type* × *Wh-Word* × *Group* was approaching significance [$F(6, 90) = 2.147$, $p = .056$]. A Bonferroni test to explore the main effect of group revealed no significant differences between the G-SLI children and the grammar controls (LA1) [$p = .277$], but there were significant differences between the G-SLI group and the two vocabulary control groups [G-SLI vs. LA2: $p < .001$; G-SLI vs. LA3: $p < .001$]. There were also significant differences between LA1 and LA2 [$p < .05$] and LA1 and LA3 [$p < .05$], but not between LA2 and LA3 [$p = 1$].

To further investigate the *Error Type* × *Group* interaction we conducted two one-way ANOVAs, one for each error type. We found a significant difference between the groups in the +wh,−T error type (no *do*-support, no tense, double tense marking) [$F(3, 45) = 7.659$, $p < .001$]. A Tukey HSD post hoc test revealed significant differences between G-SLI children and the two vocabulary control groups, LA2 [$p = .001$], and LA3 [$p < .01$], but not the grammar controls (LA1) [$p = .194$]. There was no significant difference between the control groups. The one-way ANOVA for −wh,+T errors (gap-filling, no pied-piping) also revealed a significant difference between the groups [$F(3, 45) = 16.503$, $p < .001$]. A Tukey HSD post hoc test revealed significant differences between G-SLI children and the two vocabulary control groups, LA2 [$p < .001$], and LA3 [$p < .001$], but again not the grammar controls, LA1 [$p = .135$]. There was no significant difference between the two vocabulary control groups (LA2 and LA3) [$p = .959$], but in contrast to the −T errors, there was a significant difference between the grammar and the vocabulary controls, LA1 and LA2 [$p < .05$] and LA1 and LA3 [$p < .01$]. Further, planned comparisons showed that G-SLI children and the LA1 group were significantly better in rejecting ungrammatical questions with a −wh/+T error than with a +wh/−T error [G-SLI: $t(13) = 6.595$, $p < .001$; LA1: $t(10) = 2.735$, $p < .05$].

Planned comparisons following the *Error Type* × *Question Type* interaction revealed that in both subject and object questions, the groups were performing worse in +wh,−T than in −wh,+T errors [Subject: $t(48) = 2.274$, $p < .05$; Object: $t(48) = 5.419$, $p < .001$]. However, the groups were performing equally well in the Subject and Object wh-questions involving −wh,+T errors [$t(48) = -.295$, $p = .769$], but in questions involving T-to-C dependency violations, they performed worse on object than on subject questions [$t(48) = 2.376$, $p < .05$].

Planned comparisons to further investigate the interaction between *Question Type* × *Wh-Word* revealed that in *Which*-questions children performed better in subject than in object questions [$t(48) = 3.282$, $p < .01$]. Similarly, in *Who*-questions children were better in subject than in object questions and this difference was approaching significance [$t(48) = 1.882$, $p = .066$]. In contrast, for *What*-questions there was no significant difference between subject and object questions [$t(48) = -.357$, $p = .723$].

To investigate in more detail different types of errors (e.g., lack of *do-support*, lack of tense, and double tense marking) within each error type (+wh,−T and −wh,+T), we conducted separate analyses for the +wh,−T and the −wh,+T error types. Fig. 2 shows the percentage of correct rejections for each of the different +wh,−T error types.

A 3 × 4 ANOVA with the factors *+wh,−T Error Type* (no *do-support*, no tense, double tense marking) and *Group* (SLI, LA1, LA2, LA3) revealed a significant main effect of *+wh,−T Error Type* [$F(2, 90) = 3.772$, $p < .05$] indicating differences between these error types across all groups and a significant main effect of group [$F(3, 45) = 15.238$, $p < .001$]. A Tukey HSD post hoc test revealed significant differences between the G-SLI children and the two vocabulary control groups, LA2 [$p = .001$], and LA3 [$p < .001$]. The difference between the G-SLI children and the grammar controls (LA1) did not reach significance [$p = .073$]. There was, however, a significant difference between LA1 and LA2 [$p < .05$], LA1 and LA3 [$p < .05$], but no significant difference between LA2 and LA3 [$p = .95$]. Following up the main effect of *+wh,−T Error Type*, we conducted paired sample *t*-tests to investigate significant differences between the three error types. We found a significant difference between errors of *do-support* and double tense errors [$t(48) = 2.445$, $p < .05$] but the difference between errors of *do-support* and tense errors was not significant [$t(48) = 1.791$, $p = .08$], and nor was there a reliable difference between tense and double tense errors [$t(48) = -.962$, $p = .34$]. Thus, children were more likely to incorrectly accept a question with omission of *do* than a question where tense was not marked at all, or double marked.

Fig. 3 shows the percentage of correct rejections for each one of the different −wh/+T error types.

A 3 × 4 ANOVA with the factors *−wh,+T Error Type* (gap-filling, gap-filling involving *which-one*, no pied-piping) and *Group* (SLI, LA1, LA2, LA3) revealed only a significant main effect of group [$F(3, 45) = 8.144$, $p < .001$]. A Tukey HSD post hoc test

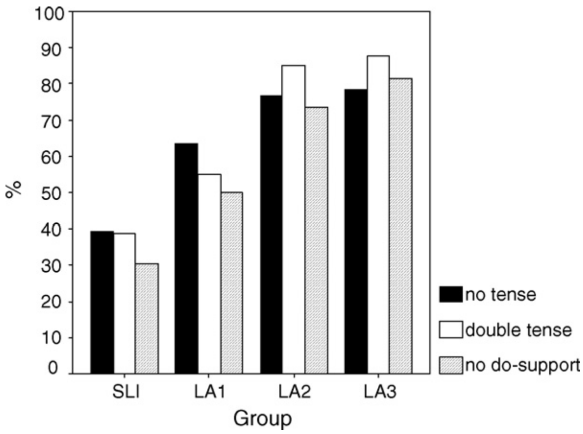

Fig. 2. Mean percentage of accurate responses in +wh,−T error type.

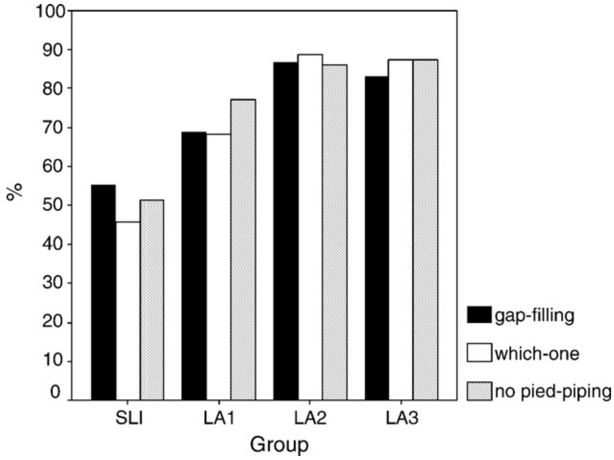

Fig. 3. Mean percentage of accurate responses in −wh,+T error type.

revealed significant differences between the G-SLI children and the two vocabulary control groups, LA2 [p = .001], and LA3 [p = .001]. There were no reliable differences between the G-SLI children and the grammar controls [p = .097], or between the LA groups [LA1 vs. LA2: p = .31; LA1 vs. LA3: p = .38; LA2 vs. LA3: p = .99].

3.3. Individual participant analysis

A final set of analyses was carried out to ascertain how many children in each group were at an above chance level in judging the −wh and −T errors (see Table 6). The criterion was set according to a .5 probability of getting one item correct by chance and on the cumulative binomial probability over the number of items for −wh and −T. The above chance level criterion was set at $p < .05$.

It can be seen from Table 6 that 11 (79%) of the 14 children with G-SLI failed to reach the above chance criterion for the −wh errors, and none reached the criterion for the −T errors. Of the 3 children (21%) reaching criterion for wh-trace dependencies, none were close to ceiling, whereas we would clearly expect teenaged individuals to be so. In contrast to the

Table 6
Percentage (number) of children passing or failing the criterion for above chance grammaticality judgements of −wh and −T errors.

Error type	G-SLI		LA1		LA2		LA3	
	Pass	Fail	Pass	Fail	Pass	Fail	Pass	Fail
−wh	21.4 (3)	78.6 (11)	54.5 (6)	45.5 (5)	83.3 (10)	16.7 (2)	83.3 (10)	16.7 (2)
−T	0 (0)	100 (14)	36.4 (4)	63.6 (7)	91.7 (11)	8.3 (1)	83.3 (10)	16.7 (2)

participants with G-SLI, the majority of the LA2 and LA3 controls and approximately half of the youngest LA1 controls correctly rejected questions with −wh and −T errors. These data indicate that, even when compared with much younger children, most (almost 80%) of the G-SLI participants exhibited significant impairment in wh-trace dependencies, and all of them exhibited significant impairment in T-to-C dependencies.

4. Discussion

This study aimed to investigate input processes and representations of children with G-SLI. It assessed whether G-SLI children could make judgements about matrix subject and object questions that were either correct, contained semantic errors, or contained one of two syntactic dependency related errors (wh-trace or T-to-C), and compared their performance with younger typically developing (TD) children matched on tests tapping grammar or vocabulary.

The results revealed a consistent pattern. First, the children with G-SLI, like the TD children, correctly accepted grammatical questions and correctly rejected semantically anomalous questions around 85% of the time. The four groups' responses for these judgements did not differ from one another. In contrast, for the syntactic dependency errors (−wh or −T) the judgements of the children with G-SLI were worse than all three control groups. This difference was consistently significant between the G-SLI and the two vocabulary control groups, although not between the G-SLI and the grammar matched controls.

The group analysis revealed that the pattern of performance across the conditions was generally similar for the TD children and the children with G-SLI. The groups' judgements were worse on the −T errors than the −wh errors. Further, the results revealed, that for −wh errors, the groups' performance on subject and object questions did not differ, and nor were there any differences between the various types of errors (e.g., gap-filling vs. no pied-piping). In contrast, −T error judgements were worse for object than subject questions. This difference can be accounted for by the "no-do-support" errors which only occur in object questions. Generally, judgements were significantly worse for these errors than no-tense marking or double tense marking. In other words, *Who Homer Simpson tickle?* was more likely to be rejected as incorrect than *Who Homer Simpson tickled?* for all the groups. Thus, children were more likely to accept ungrammatical T-to-C errors, providing tense was marked in the sentence, than if there was no tense marking or double tense marking.

The individual analysis provided further insight into the children's representations of wh-questions and syntactic dependency errors. Whereas nearly all of the older (7–8-year-old) TD vocabulary controls and approximately half of the younger (5–6-year-old) TD (LA1) controls reached the criterion for both the −wh and −T errors, only 3 (21%) of our largely teenaged G-SLI participants reached the criterion for −wh errors and none for −T errors. In other words, most of our teenagers and young adult G-SLI participants were performing at chance on all of their judgements involving syntactic dependencies in questions.

We will now discuss the implications of the findings with respect to the various hypotheses of SLI and typical development. First, the results further indicate that G-SLI children have a significant deficit in both wh-trace and T-to-C movement in that they were unable to correctly reject questions with such errors.[5] Their significantly impaired performance with syntactic dependency violations sharply contrasts with their very high performance when rejecting questions with semantic violations and accepting grammatical questions. These results concur with previous findings for young children and children with SLI: that is, their grammar is "too broad" in that they accept a wider set of sentences than are grammatical for their language. With development and learning of syntactic knowledge, or in relation to this study, development of syntactic dependencies, the syntax narrows down the possible interpretations and or productions that are possible in the grammar that the child is learning for his/her language. This pattern of results may also reflect the typically found "yes bias" when children do not know the correct answer. In either case, the results provide an additional piece of evidence showing non-adult grammar in G-SLI participants, even into young adulthood. These findings further support several current views of SLI (Friedmann and Novogrodsky, 2007; Hamann, 2006; Jakubowicz and Strik, 2008; van der Lely, 2005; van der Lely and Battell, 2003). The results indicate that decreasing the processing load in a task where only input and storage but not output is required does not reveal normal representations of wh-questions in G-SLI. This finding can be contrasted with investigations of adults with aphasia, where good performance has been reported for grammaticality judgements (Tyler, 1992), suggesting that the deficit lies in processes outside the syntactic system.

The results from this study support our previous findings investigating wh-questions in children with G-SLI. First they indicate that the −wh and −T errors made by G-SLI in elicitation tasks are not merely a result of the lead-in question or some other methodological factor. Here we use a quite different methodology and find that children accept as grammatical exactly the same type of errors that they produce in elicitation tasks (van der Lely and Battell, 2003). The data also support and extend our on-line cross-modal priming study (Marinis and van der Lely, 2007); thus not only do G-SLI participants not reactivate the trace in a gap position, but their grammar does not prevent them filling the internal verb argument position with an indefinite DP. Whereas resumptive pronouns occur in other languages, indefinite resumptives do not occur in any language as far as we are aware. Clearly, a simple auditory input processing account which claims that SLI children's

[5] An anonymous reviewer suggested that our data could instead be accommodated by a much narrower hypothesis, namely that the deficit lies in difficulties in the acquisition of *do*-insertion. This is an interesting thought, and one that we have considered seriously, but it does not fully account for our data. Wh-trace dependency itself is impaired (van der Lely and Battell, 2003); for example, children with G-SLI produce gap-filling errors in subject questions, such as *which door did it creaked*, which cannot be explained solely by a deficit in *do*-insertion.

problems are related to perceiving non-salient forms (Leonard, 1998; Montgomery, 2004) cannot account for the acceptance of wh-trace errors such as gap filling. If poor working memory or limited processing resources were playing a major role in G-SLI children's performance, then we would expect the longer chain between the wh-word and trace in object questions to have caused increased difficulties. Phillips et al. (2005) showed that a particular neural correlate (a Long Left Anterior Negativity—Long LAN) was elicited upon hearing a wh-word and remained active over several seconds until the trace position. This neural correlate is thought to represent a memory trace of the wh-word. Thus, the lack of any significant differences between subject and object questions for the −wh errors militates against working memory playing a major role in accounting for G-SLI, contra to the general processing limitation hypotheses (Leonard et al., 2000; Montgomery, 2004; Stavrakaki, 2006). This is not to discount, however, that syntax-specific mechanisms, memory traces and consequent processes that underlie the syntactic dependencies are impaired (cf. Jakubowicz and Strik, 2008). Whilst this study was not designed to tease apart these potential underlying sources of impairment in syntactic-dependencies, further investigation of these factors is warranted.

An alternative interpretation for our results is that they result from problems with knowledge of intonation patterns, as similar utterances occur in exclamations, e.g., *What a nice picture you made!* We think that this is unlikely for at least two reasons. Firstly, we have found that children with SLI show good knowledge of prosodic structure at the sentence level, showing similar ability to vocabulary matched controls (Marshall et al., 2009). Secondly, exclamations have a completely different grammar and meaning beyond the prosody.[6]

In contrast to −wh errors, −T errors were significantly worse for object than subject questions. This was found to be due the *do*-support omission errors. Such errors require not only (the less costly) covert V-to-T movement, but also T-to-C movement (Rizzi, 1990). This finding is consistent with the CGC and DCH hypotheses, where an increasing number of movement operations are predicted to cause more errors (Jakubowicz and Strik, 2008; van der Lely, 1998; van der Lely and Battell, 2003).

The additional finding that judgements of -T errors were generally lower than −wh error judgements supports previous findings on the order of acquisition of wh- and V-to-C movement. Friedmann and Lavi found, using an elicited production task, that young children acquired V-to-C movement later, and not until 4 or 5 years (Friedmann and Lavi, 2006). Furthermore, the order of acquisition is fixed, such that children do not acquire V-to-C movement before wh-movement. Our findings support this acquisition order and allow us to make further predictions with respect to G-SLI and SLI grammar more generally. Although wh-trace and T-to-C dependencies could reflect different grammatical operations and consequently syntactic processes, for reasons of parsimony we elect to test the simpler hypothesis; that is, in G-SLI they reflect a similar deficit in syntactic dependency operations. It would be interesting to know if these errors dissociate in either typical or atypical development. Such evidence would indicate that indeed these operations are qualitatively different. But as yet, we do not know of any such evidence. It is also possible to speculate that this acquisition phenomenon is due to an economy of structure (cf. Rizzi's truncation hypothesis), but this study does not directly test this hypothesis. However, the data warrant further investigation of Rizzi's proposal from an SLI perspective.

5. Conclusions

The findings from this study do not support the hypothesis that general processing resources or working memory load outside the syntactic computational system impact on performance in children with G-SLI (Montgomery, 2004). Instead, representations or operations underlying the syntactic dependencies are impaired, but the nature of these has yet to be fully investigated (Jakubowicz and Strik, 2008; Stavrakaki, 2002, 2006). We do not consider that the pattern of results can be construed as a "delay" in normal language acquisition. Many of the participants with G-SLI were in their teenage years and were still far from obtaining a level of competence that would be expected of 6–8-year-olds in the core aspects of grammar that characterise G-SLI. The fact that young children of 3–5 might occasionally produce such errors could also be because of different reasons outside the grammar and reflect the very different cognitive resources due to their age that they bring to a task. Thus, we consider that the results provide support for the view that a core deficit in G-SLI, and potentially in other SLI children with syntactic impairments, is in the computational syntactic system proper (Friedmann and Novogrodsky, 2007; van der Lely, 2005), consistent with the hypothesis that the representations/mechanisms underlying syntactic dependencies are impaired in G-SLI grammar (van der Lely and Marshall, 2010). We found a chance level of performance in the majority of G-SLI children, who are unable to consistently compute the syntactic dependencies between a moved element and its trace or the complementizer and tense functional node. This finding builds on and extents the broad range of grammatical abilities that have been previously investigated and is consistent with the CGC hypothesis (Marshall and van der Lely, 2006; Marshall et al., 2007; van der Lely, 2005; van der Lely and Marshall, 2010). Furthermore, investigation of wh-questions using different methodologies – elicited production (van der Lely and Battell, 2003), cross-modal priming (Marinis and van der Lely, 2007) and ERPs (Fonteneau and van der Lely, 2008) – converge on the same interpretation. Further studies are required to explore whether and how domain-specific memory and/or mechanisms and their processes underlying syntactic dependencies at the clause level are impaired.

[6] For example, e.g., *What a nice picture you made!* vs. **What a nice picture did you make!* and cf. *which/what nice picture did you make?* vs. * *which/what nice picture you made*? We thank David Adger for alerting us to these examples.

Acknowledgements

We would like thank the children who participated in this study and the speech and language therapists, teachers and parents of Moor House School (Hurst Green), Dawn House School (Nottingham) and Picklenash and Newent Primary School, for their help and co-operation. Thanks to Anastasia Archonti and Theo Marinis for additional analysis. We are grateful for insightful comments from the anonymous reviewers, Teresa Guasti, David Adger and the two editors of this special issue. Heather van der Lely also expresses her personal thanks to Celia Jakubowicz for the many long, thoughtful and detailed discussions about SLI, linguistic theory and life over more than 20 years. Her insight, dedication, energy and contribution were inspirational.

The research was supported by a Wellcome Trust University Award grant (Grant No. 063713), and a Leverhulme Visiting Fellowship to Heather van der Lely, which are gratefully acknowledged.

Appendix A

A.1. Game 1

These are three Aliens from another galaxy. They all want to learn how to communicate with earthlings and want you to help them. Their first mission is to learn how to ask questions. They will take it in turns to ask you a question. I want you to listen carefully and decide if they asked their question correctly. If you think the question sounds right, then you give the Alien a star. If you think the question sounds wrong, then you give the Alien a spider. The Alien who has the most stars at the end will be the winner. First we are going to have a practice, and I will help you.

A.2. Game 2

These are the three Spice Girls from different countries. This is French Spice, this is Chinese Spice and this is Eskimo Spice. They all want to learn English so that they can sing for you. But first they want you to help them learn how to ask questions. They will take it in turns to ask you a question. I want you to listen carefully and decide if they asked their question correctly. If you think the question sounds right, then you give the Spice Girl a star. If you think the question sounds wrong, then you give the Spice Girl a spider. The Spice Girl who has the most stars at the end will be the winner.

A.3. Game 3

So now this is the final! The Alien and the Spice Girl are going to ask you questions in turn. As before I want you to listen carefully and decide if their question sounds correct. If you think it is right then give them a star, and if you think it is wrong then give them a spider.

References

Adani, F., van der Lely, H., Forgiarini, M. & Guasti, M-T., in press. Grammatical feature dissimilarities make relative clauses easier: a comprehension study with Italian children. Lingua, doi:10.1016/j.lingua.2010.03.018.

Bishop, D.V.M., 1989. Test for Reception of Grammar (TROG). University of Manchester, Age and Cognitive Performance Research Centre.

Bishop, D.V.M., Adams, C.V., Norbury, C.F., 2006. Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins. Genes, Brain and Behaviour 5, 158–169.

Chomsky, N., 1999. Derivation by Phase. MIT Press, Cambridge, Mass.

Clahsen, H., Bartke, S., Goellner, S., 1997. Formal features in impaired grammars: a comparison of English and German SLI children. Journal of Neurolinguistics 10, 151–171.

Clegg, J., et al., 2005. Developmental language disorders: a follow-up in later adult life. Cognitive, language and psychosocial outcomes. Journal of Child Psychology and Psychiatry 46, 128–149.

Conti-Ramsden, G., Durkin, K., 2007. Phonological short-term memory, language and literacy: developmental relationships in early adolescence in young people with SLI. Journal of Child Psychology and Psychiatry 48, 147–156.

Crain, S., Thornton, R., 1998. Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics. MIT Press, Cambridge, MA.

Davies, L., 2002. Specific language impairment as principle conflict: evidence from negation. Lingua 112, 281–300.

de Villiers, J.G., 1991. Why questions? In: Maxfield, T.L., Plunkett, B. (Eds.), UMOP Special Edition: Papers in the Acquisition of Wh. GLSA Publications, Amherst, MA, pp. 155–173.

de Villiers, J.G., Roeper, T., 1995. Barriers, binding, and acquisition of the DP–NP distinction. Language Acquisition 4, 73–104.

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

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

Dunn, L.M., et al., 1982. British Picture Vocabulary Scale. NFER-Nelson, Windsor.

Fisher, S.E., Lai, C.S., Monaco, A.P., 2003. Deciphering the genetic basis of speech and language disorders. Annu. Rev. Neurosci. 8, 8.

Fonteneau, E., van der Lely, H.K.J., 2008. Electrical brain responses in language-impaired children reveal grammar-specific deficits. PLoS ONE 3 (3), e1832, doi:10.1371/journal.pone.0001832.

Friedmann, N., Lavi, H., 2006. On the order of acquisition of A-movement, Wh-movement and V–C movement. In: Belletti, A., et al. (Eds.), Language Acquisition and Development. Cambridge Scholars Press, Newcastle, pp. 211–217.
Friedmann, N., Novogrodsky, R., 2004. The acquisition of relative clause comprehension in Hebrew: a study of SLI and normal development. Journal of Child Language 31, 661–681.
Friedmann, N., Novogrodsky, R., 2007. Is the movement deficit in syntactic SLI related to traces or to thematic role transfer? Brain and Language 101, 50–63.
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.
Gallon, N., Harris, J., van der Lely, H.K.J., 2007. Non-word repetition: an investigation of phonological complexity in children with Grammatical SLI. Clinical Linguistics and Phonetics 21, 435–455.
Gathercole, S., 2006. Complexities and constraints in nonword repetition and word learning. Applied Psycholinguistics 27, 599–613.
Hamann, C., 2006. Speculations about early syntax: the production of wh-questions by normally developing French children and French children with SLI. Catalan Journal of Linguistics 5, 143–189.
Hamann, C., Penner, Z., Lindner, K., 1998. German impaired grammar: the clause structure revisited. Language Acquisition 7, 193–245.
Hansson, K., Nettelbladt, U., 2006. Wh-questions in Swedish children with SLI. International Journal of Speech-Language Pathology 8, 376–383.
Jakubowicz, C., 2011. Measuring derivational complexity: new evidence from typically-developing and SLI learners of L1 French. Lingua 121, 339–351.
Jakubowicz, C., Strik, N., 2008. Scope-marking strategies in the acquisition of long-distance wh-questions in French and Dutch. Language and Speech 51, 101–132.
Jakubowicz, C., et al., 1998. Determiners and clitic pronouns in French-speaking children with SLI. Language Acquisition 7, 113–160.
Kirk, S.A., McCarthy, J.J., Kirk, W.D., 1968. Illinois Test of Psycholinguistic Abilities. Illinois University Press, Urbana.
Leonard, L.B., 1998. Children with Specific Language Impairment. MIT Press, Cambridge, MA.
Leonard, L.B., et al., 2000. Tense and finiteness in the speech of children with specific language impairment acquiring Hebrew. International Journal of Language and Communication Disorders 35, 319–335.
Manzini, M.R., 1992. Locality: A Theory and Some of Its Empirical Consequences (Linguistic Inquiry Monographs 19). MIT Press, Cambridge, MA.
Manzini, M.R., 1995. From merge and move to form dependency. UCL Working Papers in Linguistics 7, 323–345.
Marcus, G.F., Fisher, S.E., 2003. FOXP2 in focus: what can genes tell us about speech and language? Trends in Cognitive Sciences 7, 257–262.
Marinis, T., van der Lely, H.K.J., 2007. On-line processing of questions in children with G-SLI and typically developing children. International Journal of Language and Communication Disorders 42, 557–582.
Marshall, C.R., van der Lely, H.K.J., 2006. A challenge to current models of past tense inflection: the impact of phonotactics. Cognition 100, 302–320.
Marshall, C.R., van der Lely, H.K.J., 2007. The impact of phonological complexity on past tense inflection in Grammatical-SLI. International Journal of Speech-Language Pathology 9, 191–203.
Marshall, C.R., Marinis, T., van der Lely, H.K.J., 2007. Passive verb morphology: the effect of phonotactics on passive comprehension in typically developing and Grammatical-SLI children. Lingua 117, 1434–1447.
Marshall, C.R., et al., 2009. The link between prosody and language skills in children with SLI and/or dyslexia. International Journal of Language and Communication Disorders 44, 466–488.
Montgomery, J., 2004. Sentence comprehension in children with specific language impairment: effects of input rate and phonological working memory. International Journal of Language and Communication Disorders 39, 115–133.
Pesetsky, D., 1987. Wh-in-situ: movement and unselective binding. In: Reuland, E., ter Meulen, A. (Eds.), The Representation of (In)definiteness. MIT Press, Cambridge, MA, pp. 98–129.
Phillips, C., Kazanina, N., Abada, S.H., 2005. ERP effects of the processing of syntactic long-distance dependencies. Cognitive Brain Research 22, 407–428.
Rice, M., Wexler, K., 1996a. A phenotype of specific language impairment: extended optional infinitives. In: Rice, M. (Ed.), Towards a Genetics of Language. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 215–237.
Rice, M., Wexler, K., 1996b. Toward tense as a clinical marker of specific language impairment in English-speaking children. Journal of Speech and Hearing Research 39, 1239–1257.
Rizzi, L., 1990. Relativized Minimality. MIT Press, Cambridge, MA.
Rizzi, L., 2000. Remarks on early null subjects. In: Friedemann, M.-A., Rizzi, L. (Eds.), The Acquisition of Syntax. Longman, London, pp. 269–292.
Schulz, P., Roeper, T., 2011. Acquisition of exhaustivity in wh-questions: a semantic dimension of SLI? Lingua 121, 383–407.
Stavrakaki, S., 2001. Comprehension of reversible relative clauses in specifically language impaired and normally developing Greek children. Brain and Language 77, 419–431.
Stavrakaki, S., 2002. Sentence comprehension in Greek SLI children. In: Windsor, F., Kelly, M.L., Hewlett, N. (Eds.), Investigations in Clinical Linguistics and Phonetics. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 57–72.
Stavrakaki, S., 2006. Developmental perspectives on specific language impairment: evidence from the production of wh-questions by Greek SLI children over time. International Journal of Speech-Language Pathology 8, 384–396.
Stromswold, K., 1995. The acquisition of subject and object wh-questions. Language Acquisition 4, 5–48.
Thornton, R., 1990. Adventures in long-distance moving: the acquisition of complex wh-questions. Doctoral Dissertation, University of Connecticut, Storrs.
Thornton, R., 1995. Referentiality and wh-movement in child English: Juvenile D-linkuency. Language Acquisition 4, 139–175.
Tyler, L.K., 1992. Spoken Language Comprehension: An Experimental Approach to Disordered and Normal Processing. MIT Press, Cambridge, MA.
van der Lely, H.K.J., 1996a. Specifically language impaired and normally developing children: verbal passive vs. adjectival passive sentence interpretation. Lingua 98, 243–272.
van der Lely, H.K.J., 1996b. The Test of Active and Passive Sentences (TAPS). Centre for Developmental Language Disorders and Cognitive Neuroscience, University College London.
van der Lely, H.K.J., 1998. SLI in children: movement, economy and deficits in the computational syntactic system. Language Acquisition 72, 161–192.
van der Lely, H.K.J., 1999. Verb Agreement and Tense Test (VATT). Centre for Developmental Language Disorders and Cognitive Neuroscience, University College London.
van der Lely, H.K.J., 2004. Evidence for and implications of a domain-specific grammatical deficit. In: Jenkins, L. (Ed.), Variations and Universals in Biolinguistics. Elsevier, Oxford, pp. 117–144.
van der Lely, H.K.J., 2005. Domain-specific cognitive systems: insight from Grammatical-SLI. Trends in Cognitive Sciences 9, 53–59.
van der Lely, H.K.J., Battell, J., 2003. Wh-movement in children with grammatical SLI: a test of the RDDR hypothesis. Language 79, 153–181.
van der Lely, H.K.J., Christian, V., 2000. Lexical word formation in children with grammatical SLI: a grammar-specific versus an input-processing deficit? Cognition 75, 33–63.
van der Lely, H.K.J., Marshall, C.R., 2010. Grammatical-specific language impairment: a window onto domain-specificity. In: Guendouzi, J., Locke, F., Williams, M.J. (Eds.), The Handbook of Psycholinguistic and Cognitive Processes: Perspectives in Communication Disorders. Psychology Press, Hove, pp. 403–419.
van der Lely, H.K.J., Stollwerck, L., 1996. A grammatical-specific language impairment in children: an autosomal dominant inheritance? Brain and Language 52, 484–504.
van der Lely, H.K.J., Stollwerck, L., 1997. Binding theory and grammatical specific language impairment in children. Cognition 62, 245–290.
van der Lely, H.K.J., Ullman, M., 2001. Past tense morphology in specifically language impaired children and normally developing children. Language and Cognitive Processes 16, 113–336.
van der Lely, H.K.J., Rosen, S., McClelland, A., 1998. Evidence for a grammar-specific deficit in children. Current Biology 8, 1253–1258.
van der Lely, H.K.J., Rosen, S., Adlard, A., 2004. Grammatical language impairment and the specificity of cognitive domains: relations between auditory and language abilities. Cognition 94, 167–183.

Vernes, S.C., et al., 2008. A functional genetic link between distinct developmental language disorders. New England Journal of Medicine 359, 2337–2345.
Weissenborn, J., Roeper, T., de Villiers, J.G., 1995. Wh-acquisition in French and German: connections between case, Q-features and unique triggers. Recherches Linguistiques 24, 125–155.
Wong, A.M., et al., 2004. Questions without movement: a study of Cantonese-speaking children with and without specific language impairment. Journal of Speech, Language, and Hearing Research 47, 1440–1453.