RASK 49
SPRING 2019

Grammatical characteristics of narratives told by Danish children with developmental language disorder

RIKKE VANG CHRISTENSEN

Københavns Universitet, Institut for Nordiske Studier og Sprogvidenskab (NorS)

ABSTRACT
The aim of the study was to explore the structural characteristics, including the percentage of grammatical utterances (PGU), of a narrative generation task based on the wordless picture book, 'A Boy, a Dog and a Frog'. Thirty-eight Danish children aged 6;6 (years;months) to 13;2 participated. Nineteen participants with developmental language disorder (DLD) were individually matched to a typically developing (TD) child for age and gender.
Compared to their TD peers, the children with DLD produced shorter utterances and were less accurate with respect to verb morphology and word order. PGU showed good potential for identifying the participants as DLD or TD.

KEYWORDS
narrative skills, assessment, grammatical accuracy, Danish, developmental language disorder (DLD)

1. Introduction

1.1 Developmental language disorder

Developmental language disorder (DLD) refers to children who have language problems that constitute obstacles for participation in everyday activities and are not likely to resolve spontaneously (Bishop, Snowling, Thompson, Greenhalgh & the CATALISE-2 consortium 2017). The disorder is not associated with any known biomedical condition (Bishop et al. 2017). DLD thus replaces specific language impairment (SLI) as the term for a common developmental language disorder with an estimated prevalence of 7 % (Norbury et al. 2016; Tomblin et al. 1997).

DLD is characterized by deficits in aspects of language production, comprehension or use, but no specific language symptoms must be present in DLD. Symptoms are to some extent dynamic and language specific, but morphosyntax seems to be an area of particular difficulty (see Leonard 2014 for extensive review). DLD may also have long-lasting effects on academic

rikkec@hum.ku.dk

achievements and employment (e.g. Elbro, Dalby & Maarbjerg 2011) and social and emotional wellbeing (e.g. St. Clair, Pickles, Durkin & Conti-Ramsden 2011). There is thus good reason to study DLD in the given language to provide children and adolescents with high quality clinical services in the attempt to counter negative consequences of the disorder.

To make well-informed decisions, clinicians must draw on a variety of data, e.g. standardized tests and the childs' case history (Friberg 2010). When assessing characteristics beyond single words or sentences, discourse samples constitute an important supplement to standardized tests (e.g. Eisenberg & Guo 2015) since they offer a wealth of possibilities for analyses (see section 1.3.).

1.2 DLD in Danish

Danish DLD has not been the subject of much scientific scrutiny but recent studies revealed that verb past tense verb inflection (Lum & Bleses 2012; Vang Christensen & Hansson 2012) and relative clauses (Jensen de Lopez, Sundahl Olsen & Chondrigianni 2013) are challenging for Danish children with DLD.

Danish is a Germanic language, very similar to Norwegian and Swedish, with inflection of verbs, nouns and adjectives and with case-marked personal pronouns. Danish verbs are marked for tense (finiteness), but not for person or number. The basic word order is subject-verb-object (SVO), but since Danish is a verb-second (V2) language, subject-verb inversion occurs in contexts of topicalization, as in Example 1.

1. *Igår køb-te mor-en æble-r* ('yesterday buy-$_{PAST}$ mother-$_{COMMON,DEF}$ apple-$_{PLUR}$')

Word order difficulties of this type have been found in Swedish DLD (Hansson, Nettelbladt & Leonard 2000) and based on data from a sentence repetition task, Vang Christensen (2019a) found that difficulties with inversion also occur in Danish DLD (see Example 2).

2. a. Target: *Igår vent-ede*$_{VERB}$ *kone-n*$_{SUBJECT}$ *på min søde bror* ('yesterday wait-$_{PAST}$ wom an-$_{COMMON,DEF}$ for my$_{COMMON}$ nice brother$_{COMMON}$')
   b. response from child with $_{DLD}$: **Igår min kone*$_{SUBJECT}$ *vent-e*$_{VERB}$ *på den lille dreng* ('yesterday my wife wait-$_{INF}$ for the$_{COMMON,DEF}$ little boy$_{COMMON}$').

Danish generally has more elaborate inflectional systems than English. For instance, Danish nouns have arbitrary gender, common or neuter, and definite, singular nouns are inflected for gender: -(*e*)*n*$_{COMMON,}$ (e.g. *bold-en* 'ball-$_{COMMON,DEF}$'), and –(*e*)*t*$_{NEUTER}$ (e.g. *bord-et* 'table-$_{NEUTER,DEF}$'). Determiners must agree with the noun in definiteness and gender as illustrated in the final noun phrase (NP) in Example 3 where a child with DLD repeats a sentence using the common, definite article with neuter gender *vand* ('water').

3. a. Target: *Hopp-ede Klaus på hoved-et*$_{\text{NEUTER}}$ *i det*$_{\text{NEUTER, DEFINITE}}$ *kold-e vand*$_{\text{NEUTER}}$*?* ('Jump-$_{\text{PAST}}$ Klaus on head-$_{\text{NEUTER,DEF}}$ in the cold-$_{\text{DEF}}$ water?')
   b. Repetition: **Kaus hopp-e i den*$_{\text{COMMON,DEF}}$ *dyb-e vand*$_{\text{NEUTER}}$ ('*Kaus jump-$_{\text{INF}}$ in the deep-$_{\text{DEF}}$ water').

Difficulties with neuter gender determiners have also been found in Swedish DLD (see Hansson, Nettelbladt & Leonard 2003).

1.3 Narratives and narrative skills in DLD

Narratives play a major role in human communication and are important for educational and social success (see Boudreau 2008 for review). A narrative is a monologic type of discourse, which can be defined as a sequence of utterances about one or more events that unfold over time and where the plans and goals of an agent are included (Norbury, Gemmell & Paul 2014: 486). Narrative production is dependent on a range of skills, from language skills, e.g. vocabulary, morphosyntax and pragmatics, over broader cognitive skills such as attention and working memory to social skills, e.g. taking the perspective of the listener or the story characters (e.g. Johnston 2006). Simultaneous processing at several levels is thus necessary for producing a well-formed narrative. Norbury & Bishop (2003) pointed to the central influence of core language skills, e.g. morphosyntax, on the quality and characteristics of narratives.

Wetherell, Botting and Conti-Ramsden (2007) pointed to 'narrative' as an umbrella term covering a range of different formal and informal tasks. A major distinction is between narrative retelling tasks and generation tasks. One frequently used format is narrative generation from a picture book (e.g. Engberg-Pedersen & Vang Christensen 2017). Here children construct a narrative that fits the visual input from the pictures, without being provided with a language model.

In a study of 117 British children with DLD, Botting, Faragher, Simkin, Knox and Conti-Ramsden (2001) found that a strong predictor of language disorder at age 11 was narrative performance – assessed with a retelling task – at age 7. It thus seems highly relevant that narrative comprehension and production were included in the assessment battery in the epidemiological studies of DLD (Norbury et al. 2016; Tomblin et al. 1997) and the predictive potential further points to the relevance of including narratives in clinical assessment.

Narrative tasks are relatively naturalistic yet sufficiently structured to allow for comparisons between individuals or groups (Botting 2002). They have been used extensively in studies of children with DLD during the last decades (e.g. Duinmeijer, de Jong & Scheper 2012; Fey et al. 2004; Norbury & Bishop 2003). With respect to language structure, narrative tasks are used both for broad characterization of language production, e.g. average utterance length or level of morphosyntactic accuracy and for analyses of particular features such as use of subordinate clauses or accuracy of tense marking.

'Mean length of utterance in morphemes' (MLU_m) is a common measure of morphosyntactic complexity in early language development (e.g. DeThorne, Johnson & Loeb 2005). Frizelle, Thomson, McDonald and Bishop (2018) argued that MLU is a relevant measure beyond the ear-

liest stages of language development as there is evidence that the utterance length – usually computed in number of words, $MLU_{w,}$ rather than morphemes in advanced language users – continues to increase into adolescence and adulthood (e.g. Nippold, Hesketh, Duthie & Mansfield 2005). However, Frizelle et al. (2018) emphasized that MLU_w does not adequately capture syntactic complexity since a number of means can increase utterance length. One such means, which will also be employed in the present study, is the use of finite subordinate clauses as in Example 4 from a 13-year-old boy with DLD who participated in the present study. He used subordination, but with inaccurate word order since he uttered the finite verb before rather than after the adverb (compare Example 4a to 4b).

4. a. **de prøv-er at led-e efter en frø fordi de vil gerne leg-e med den og se hvad den kan* ('they try-$_{PRES}$ to search-$_{INF}$ for a frog because they would like play-$_{INF}$ with it and see what it can')
   b. Accurate alternative: … *fordi*$_{CONJUNCTION}$ *de*$_{SUBJECT}$ *gerne*$_{ADVERB}$ *vil*$_{VERB}$ *leg-e med den* … ('… because they like would play-$_{INF}$ with it …')

Frizelle et al. (2018) investigated utterance complexity, specifically the use of subordinate clauses, in narratives collected from British participants aged from 4 years to adulthood. Overall, MLU_w and syntactic complexity increased with age. Surprisingly, no clear changes in MLU_w or use of subordinate clauses were seen from 7 to 12 years of age. This result could reflect that unless children in this age range are specifically encouraged to use their full linguistic competence they may not do so (Frizelle et al. 2018).

Eisenberg & Guo (2013) used 'percentage of grammatical utterances' (PGU) as a broad measure of children's morphosyntactic skills. They calculated grammatical utterances and utterances containing one or more errors, e.g. omissions of obligatory arguments such as subject or tense marking errors of verbs. Based on data from a picture description task, these researchers showed that PGU reliably differentiated 3-year-olds with and without language disorders. All participants with language disorder ($n = 17$) were accurately identified and the overall identification accuracy was 94 % (32 of 34 participants) (Eisenberg & Guo 2013). The PGU measure is thus a candidate for a clinical marker of DLD in young children (see also Eisenberg & Guo 2015) – and possibly in 6 to 8-year-old children too (Guo & Schneider 2016).

For a measure to work as a clinical marker, it must have high levels of 'sensitivity', i.e. the proportion of children with language disorder who are correctly identified as language disordered with the task, and 'specificity', i.e. the proportion of children with typical development who accurately attain a 'non-disordered' score. Sensitivity and specificity levels from 80 % are considered acceptable, whereas values above 90 % are high (Plante & Vance 1994).

Given the demands on linguistic and other cognitive skills posed by elicited narratives, these tasks seem suitable for revealing grammatical weaknesses in children with DLD, even when they have developed sufficiently for managing everyday conversation quite inconspicuously. Fey et al. (2004) analyzed narratives generated from short picture sequences by 2nd and 4th graders with and without DLD. Children with DLD generated stories with shorter MLU and

lower PGU levels. Duinmeijer et al. (2012) found similar narrative difficulties in Dutch children with DLD aged 6;1 to 9;9 relative to typically developing (TD) children in narrative retelling and generation tasks. Participants with DLD produced more complex utterances, i.e. utterances containing more than one finite verb, in the retelling task compared to the generation task. On this basis, the authors argued that narrative generation best reflects the child's linguistic level. The results of this study illustrate how the characteristics of the narrative task may influence results (see also Frizelle et al. 2018), particularly in children with DLD whose performance is hampered by poor language skills. Finally, Wetherell et al. (2007) investigated syntactic complexity and syntactic errors in 13 to 15-year-olds with and without DLD in narrative tasks. These adolescents with DLD did not differ significantly from the controls with respect to the number of clauses produced or their use of complex utterances (e.g. subordinate clauses), but they produced significantly more morphosyntactic errors, e.g. omissions of obligatory elements or of morphological markers of tense or agreement. Interestingly, the participants with DLD showed a larger increase in errors from a conversational task to the more taxing generation task than their TD peers; this interaction between task and participant group possibly reflects the linguistic vulnerability of adolescents with DLD.

1.4 Aims of the study

The results of previous studies suggest that narrative tasks can be used for assessing structural language skills in DLD and for identifying weaknesses in morphosyntax, even in adolescents with DLD. The aim of the present study was therefore to examine the grammatical characteristics of the narrative 'A Boy, a Dog and a Frog' generated by Danish children with and without DLD in the age range 6 to 13 years. The study addressed the following questions:

1. Do children with DLD perform below the level of their TD peers in narrative generation with respect to (a) utterance length (MLU_w); (b) use of finite subordinate clauses (number and accuracy); (c) grammatical accuracy – measured with PGU and with calculation of specific error types (ungrammatical verb inflection, word order errors, omission of verb or of arguments)?
2. Does PGU identify Danish school-aged children with DLD or TD with acceptable levels of sensitivity and specificity?

2. Method

2.1 Participants

Data from 38 children from Greater Copenhagen was included in the present study. These children were participants in a larger project, (1) Language and Cognition – Perspectives from Impairment (LaCPI) or (2) Language Mentor[1]. In accordance with the guidelines provided by The

1 see https://nors.ku.dk/english/research/projects/communication-disorders for more information about the two projects.

Faculty of Humanities' Research Ethics Committee, all parents had given informed written consent to their child's participation and they were informed of the right to withdraw from the study at any point. Data was stored and processed in full compliance with EU data protection guidelines, GDPR.

According to parental and school reports, no participant had permanent hearing loss requiring treatment or neurological/motor/socio-emotional impairment, and there were no concerns about nonverbal cognitive skills. All participants spoke Danish as their primary language. Half of the participants (n = 19) were TD children recruited from ordinary schools. The other half (n = 19) were identified with DLD by a speech and language pathologist (SLP) independent from the study and they all received language intervention. One participant with DLD was recruited from the caseload of an SLP. All other participants with DLD were recruited from special school units for children with marked language difficulties but unimpaired cognitive skills.

For confirmation of the status as DLD or TD, the language proficiency was assessed with three language tests: (1) 'Expressive vocabulary', a standardized picture naming task developed for research purposes (84 color photos of objects such as *kiwi* and *saxophone*; Gellert & Vang Christensen 2012)[2]; (2) the Danish version of the 'Test of Reception of Grammar-2' ('TROG-2', Bishop 2010); and (3) the subtest 'Recalling Sentences' from the Scandinavian version of the Clinical Evaluation of Language Fundamentals, 4th edition (CELF-4; Semel, Wiig & Secord 2013).

The -1.25 SD cut-off is often used in DLD research (e.g. Tomblin et al. 1997). Participants with TD scored better than -1.25 SD (the 10th percentile) on all three language measures, whereas participants with DLD scored at or below -1.25 SD on at least two of the included language tests.

Finally, participants were tested with a measure of nonverbal skills, the 'Matrices' subtest of the Wechsler Non-verbal Scale of Ability (WNV, Wechsler & Naglieri 2009). All participants scored better than T-score 30 (-2 SD) (see Table 1). This confirmed that they were not intellectually impaired.

Participants with and without DLD were matched pairwise for age (+/- two months) and gender. They form four groups (see Table 1): (1) younger children with DLD (Y-DLD, n = 10, five girls and five boys) aged 6;7 to 8;11; (2) younger TD children (Y-TD, n = 10) aged 6;6 to 9;0; (3) older children with DLD (O-DLD, n = 9, three girls and six boys) aged 11;1 to 13;1; and (4) older TD children (O-TD, n = 9) aged 11;2 to 13;2.

2.2 Procedure

Participants were assessed by the author or by trained students of SLP, linguistics or psychology. The assessments took place in a separate room at the participants' school. Language production tasks were audio-recorded digitally on an Olympus LS-11 recorder for subsequent transcription and scoring (see section 2.4.).

2 The measure of productive vocabulary does not have norms. Thus, the results from approximately 130 TD children formed the basis for calculating the -1.25 SD cut-off for each grade-level. These children (including the TD participants in the present study) took part in the research project Language and Cognition – Perspectives from Impairment (see note (1)).

2.3 Narrative generation task

Narratives were elicited by means of the wordless picture book 'A Boy, a Dog, and a Frog' (Mayer 1967). First participants leafed through the book to familiarize themselves with the story. Then they told the story to the experimenter while looking at the pictures. This approach emphasized the visual support provided by the pictures, but also signaled that participants were not free to generate whatever story they pleased. The experimenter responded by back channel signals such as *oh* or *okay* on appropriate occasions, but she looked away to encourage the participant to tell a comprehensive narrative and to reduce the risk of the participant pointing to details in the pictures.

'A Boy, a Dog, and a Frog' is about a boy who goes fishing together with his dog. He sees a frog in a pond and wants to catch it. The boy attempts to catch the frog in various ways also involving the dog, without success. The boy gives up and walks home with his dog. The frog follows their footprints and reaches their home. The frog finds the boy and the dog in the bathtub. At the end, the frog joins the boy and the dog.

This frog story has not been used as widely in the research literature as 'Frog, where are you?' (Mayer 1969). It was chosen for the LaCPI project because the intentions and emotional reactions of the characters are quite clear throughout the story (see also Norbury et al. 2014). Thus, we assumed that the visual stimuli would encourage participants to talk about the mental states of the characters – and that this would prompt the use of complex syntax, e.g. subordinate clauses (Engberg-Pedersen & Vang Christensen 2017).

2.4 Transcription

Trained students transcribed the narratives according to guidelines developed in the LaCPI project and instances of doubt were resolved in dialogue with the author who subsequently went over all transcriptions while listening to the recordings of the narratives. Only few minor adjustments to the transcriptions were made at this point. Two SLP students scored the transcriptions for the relevant morphosyntactic categories (see section 2.5.) before a third SLP student re-scored the material. Finally, these re-scorings were checked by the author with a very high level of agreement (no formal calculation of interrater agreement was performed).

2.5 Scoring

The narratives were divided into 'C-units' (communication units), defined as each independent clause with its dependent clauses (e.g. Fey et al. 2004); Example 4 in section 1.3. and Examples 5 to 7 below each constitute one C-unit. Fragments (utterances without a verb) and utterances lacking obligatory arguments such as the subject (see Example 6) were counted as C-units in accordance with Eisenberg and Guo (2013, 2015) and Frizelle et al. (2018) since children with DLD are likely to produce such utterances.

5. a. **dreng*$_{\text{COMMON}}$ *kravl-e op af træ*$_{\text{NEUTER}}$*-en* ('boy climb-$_{\text{INF}}$ up of tree-$_{\text{COMMON,DEF}}$')
   b. accurate alternative: *dreng-en kravl-ede op i træ-et* ('boy-$_{\text{COMMON,DEF}}$ climb-$_{\text{PAST}}$ up in tree-$_{\text{NEUTER,DEF}}$')

6. a. **fang-et frø* ('caught-$_{\text{PARTICIPLE}}$ frog')
   b. accurate alternative: *dreng-en fang-ede frø-en* ('boy-$_{\text{COMMON,DEF}}$ catch-$_{\text{PAST}}$ frog-$_{\text{COMMON,DEF}}$')
7. a. **hun f- dreng f- fang-et til h- hund* ('boy catch-$_{\text{PARTICIPLE}}$ to dog')
   b. accurate alternative: *dreng*$_{\text{COMMON}}$*-en fangede hund*$_{\text{COMMON}}$*-en* ('boy-$_{\text{COM MON,DEF}}$ catch-$_{\text{PAST}}$ dog-$_{\text{COMMON,DEFINITE}}$')

As measures of story length, the number of C-units and the number of words were calculated. Non-narrator comments such as *skal jeg fortælle hvad han råber* ('should I tell what he yells') or *nu er den slut* ('now is it finished') were excluded from the C-unit calculation. Self-repetitions, self-corrections, false starts, fillers and incomprehensible words were excluded from the word count as illustrated by the struck through parts of Example 7.

MLU_w was included as a global measure of expressive language skills (see DeThorne et al. 2005). MLU_w was computed based on C-units by dividing the number of words by the number of C-units (see Fey et al. 2004 for same procedure).

One means of increasing utterance length and syntactic complexity is use of subordinate clauses (see Example 4). The number of finite subordinate clauses and the percentage of morphosyntactically accurate finite subordinate clauses were computed to further characterize the participants.

The PGU was calculated as a measure of morphosyntactic proficiency by dividing the number of morphosyntactically accurate C-units by the total number of C-units. C-units with one or more omissions or substitution errors of free or bound morphemes (see Examples 5 to 7) were scored as inaccurate. Inflected forms which could be interpreted as overregularizations of e.g. past tense (**følg-ede* for *fulgte* ('followed')) or common gender definiteness inflection (**menneske-n*$_{\text{COMMON,DEF}}$ for *menneske-t*$_{\text{NEUTER,DEF}}$ ('human-$_{\text{DEF}}$')) were computed as inaccurate since the distinction between overregularization and inaccurate use may be highly arbitrary. In Example 6, the subject was omitted and in Example 8 the object was omitted; such C-units were also coded as inaccurate. For further description of the scoring criteria please refer to Eisenberg and Guo (2013, 2015).

8. a. **og så dreng-en*$_{\text{SUBJECT}}$ *løft-er*$_{\text{VERB}}$ *op* ('and then boy-$_{\text{COMMON,DEF}}$ lift-$_{\text{PRES}}$ up')
   b. accurate alternative: *og så løft-er*$_{\text{VERB}}$ *dreng-en*$_{\text{SUBJECT}}$ *sig/frø-en op* ('and then lift-$_{\text{PRES}}$ boy-$_{\text{COMMON,DEF}}$ himself/frog-$_{\text{COMMON,DEF}}$ up')

Two types of word order errors were also coded as inaccurate: (1) subject-verb inversion errors in contexts of topicalization as in Example 8; and (2) errors with placement of negation or adverb as in Example 4 and Example 9.

9. a. **og hund-en er blevet også meget sur* ('and dog-$_{\text{COMMON,DEF}}$ has become also very angry')
   b. accurate alternative: *og hunden er også blevet meget sur* ('and dog-$_{\text{COMMON,DEF}}$ has also become very angry')

Finally, specific error types were calculated in order to (1) characterize the morphosyntax of the participants further, and (2) investigate the usefulness of the narrative context for revealing errors commonly found in DLD. 'Ungrammatical verb inflection', e.g. a nonfinite form such as the infinitive (see Example 5) or the participle (see Example 6) in a finite context was calculated along with the following syntactic inaccuracies: 'fragment' (omission of the verb); 'argument omission' and 'word order errors'.

2.6 Statistics

Due to the small group sizes and non-normal distribution of the data, group comparisons were conducted with the non-parametric Mann-Whitney test, with the 'exact' two-tailed option for p-values. These calculations were performed in SPSS version 25 (IBM Corp. 2017). Effect sizes (r-values) were calculated for significant group differences. Effect sizes from 0.5 (explaining from 25 % of the variation) were considered large whereas effect sizes from 0.3 were moderate (Field 2009). The r-values are reported along with Z- and p-values in the Results section. Furthermore, ROC-analyses were performed in SPSS to investigate the diagnostic potential of the PGU measure.

3. Results

3.1 Background assessment

The participants in the study were not intellectually impaired and no child scored below -1.5 SD (T-score = 35) on Matrices (see Table 1). Whereas the older groups were well matched on this measure, the Y-DLD group scored significantly lower than the Y-TD group (see Table 1). Such differences are not unusual in the research literature (e.g. Fey et al. 2004); they may reflect the non-specific nature of DLD (Bishop et al. 2017).

Both groups with DLD scored lower on the included language tests than their TD peers, with medium or large effect sizes (see Table 1). Note that only the younger groups were compared on 'Recalling Sentences', since only one participant from the O-TD group completed this task.

Comparisons of the O-DLD and Y-TD groups with respect to age and vocabulary were also conducted. The participants in the O-DLD group were significantly older than the Y-TD participants ($z = 3.68$, $p < .001$, $r = .85$), yet they scored much lower on Expressive vocabulary than their younger TD peers ($z = 3.23$, $p < .001$, $r = .74$).

TABLE 1: MEANS (*M*), STANDARD DEVIATIONS (*SD*) AND RANGE OF SCORES ON AGE, MATRICES (THE MEASURE OF NONVERBAL COGNITIVE SKILLS), AND THREE LANGUAGE TESTS FOR THE YOUNGER AND OLDER GROUPS WITH DLD OR TD

	Younger DLD (Y-DLD)				Younger TD (Y-TD)				
	n	*M*	*SD*	Range	*n*	*M*	*SD*	Range	
Age (years;-months)	10	7;10	0;8	6;7-8;11	10	7;9	0;8	6;6-9;0	Y-DLD ≈ Y-TD: $p = 0.898$
Matrices (T-score)	10	46.7	7.3	41-54	10	56.2	6.8	43-65	Y-DLD < Y-TD: $z = -2.73$, $p = 0.005$, $r = 0.610$
Expressive vocabulary (raw score, max.84)	10	38.1	9.1	21-55	10	61.9	4.3	55-69	Y-DLD < Y-TD: $z = -3.74$, $p < 0.001$, $r = 0.838$
TROG-2 (standard score)	10	73.9	13.7	54-88a	10	107.1	9.2	93-120	Y-DLD < Y-TD: $z = -3.78$, $p < 0.001$, $r = 0.846$
'Recalling Sentences', CELF-4 (scaled score)	10	3.6	2.0	1-6	10	11.2	2.0	8-15	Y-DLD < Y-TD: $z = -3.80$, $p < 0.001$, $r = 0.850$
	Older DLD (O-DLD)				**Older TD (O-TD)**				
	n	*M*	*SD*	Range	*n*	*M*	*SD*	Range	
Age (years;-months)	9	11;11	0;10	11;1-13;1	9	12;0	0;10	11;2-13;2	O-DLD ≈ O-TD: $p = 0.914$
Matrices (T-score)	9	46.9	7.3	35-56	9	48.1	6.3	38-61	O-DLD ≈ O-TD: $p = 0.985$
Expressive vocabulary (raw score, max.84)	9	48.7	8.6	37-61	9	74.2	4.5	67-79	O-DLD < O-TD: $z = -3.58$, $p < 0.001$, $r = 0.844$
TROG-2 (standard score)	9	78.6	15.1	55-101	9	103.2	7.1	91-111	O-DLD < O-TD: $z = -3.11$, $p = 0.001$, $r15 = 0.735$
'Recalling Sentences', CELF-4 (scaled score)	9	2.6	2.3	1-7	1b	-	-	12	-

a) In an instance of an extremely low standard scores on *TROG-2*, the score 54 was entered for data processing purposes since the norms do not specify scores <55.
b) Due to a heavy load of testing on the older TD children within a larger research project, only one participant in this group went through this testing.

TABLE 2: MEANS (*M*), STANDARD DEVIATIONS (*SD*) AND RANGE FOR NUMBER OF UTTERANCES, NUMBER OF WORDS, MLU_W, NUMBER AND ACCURACY PERCENTAGE OF SUBORDINATE CLAUSES AND FINALLY PGU (PERCENTAGE GRAMMATICAL UTTERANCES) FOR THE YOUNGER AND OLDER GROUPS WITH DLD OR TD

	Younger DLD (Y-DLD) (*n* = 10)			Younger TD (Y-TD) (n = 10)			
	M	**SD**	**Range**	***M***	***SD***	**Range**	
Number of C-units	36.6	11.9	22-56	30.5	6.3	16-38	Y-DLD ≈ Y-TD: *p* = 0.362
Number of words	212.7	89.9	107-372	238.9	50.3	147-319	Y-DLD ≈ Y-TD: *p* = 0.436
MLUw	5.68	0.98	3.96-7.41	7.94	1.20	6.21-9.80	Y-DLD < Y-TD: *z* = -3.25, *p* < 0.001, *r* = 0.73
Number of finite subordinate clauses	6.1	6.2	0-21	5.6	3.5	2-12	Y-DLD ≈ Y-TD: *p* = 0.836
Finite subordinate clauses – percentage accurate	68.8[a]	26.2	33.3-100	87.0	17.3	50-100	Y-DLD ≈ Y-TD: *p* = 0.146
PGU	56.1	22.3	6.8-74.1	86.0	8.6	74.2-97.1	Y-DLD < Y-TD: *z* = -3.78, *p* < 0.001, *r* = 0.85
	Older DLD (O-DLD) (*n* = 9)			**Older TD (O-TD) (*n* = 9)**			
	M	***SD***	**Range**	***M***	***SD***	**Range**	
Number of C-units	38.4	14.7	22-63	27.6	5.9	18-33	O-DLD ≈ O-TD: *p* = 0.230
Number of words	269.3	136.4	145-498	251.4	70.7	157-350	O-DLD ≈ O-TD: *p* = 1.0
MLUw	6.86	1.44	4.43-9.64	9.10	1.51	6.28-11.58	O-DLD < O-TD: *z* = -2.43, *p* = 0.014, *r* = 0.57

(a) based on the eight Y-DLD participants who produced finite subordinate clauses.

TABLE 2 CONTINUED

	Older DLD (O-DLD) (*n* = 9)			Older TD (O-TD) (*n* = 9)			
	M	***SD***	**Range**	***M***	***SD***	**Range**	
Number of finite subordinate clauses	4.7	5.2	0-14	6.8	3.5	2-14	O-DLD ≈ O-TD: $p = 0.140$
Finite subordinate clauses – percentage accurate	69.6[b]	38.2	0-100	85.6	17.2	50-100	O-DLD ≈ O-TD: $p = 0.645$
PGU	76.1	11.2	58.7-92.3	90.1	5.3	80.8-100	O-DLD < O-TD: $z = -2\text{-}78$, $p = 0.004$, $r = 0.66$

(b) based on the seven O-DLD participants who produced finite subordinate clauses.

3.2 Length and morphosyntactic accuracy of the narratives

Neither the Y-DLD nor the O-DLD group from their same-age peers in the TD-groups with respect to number of C-units or number of words (see Table 2). Furthermore, the O-DLD did not differ from the Y-TD group on either number of C-units ($p = 0.434$) or number of words ($p = 0.905$). On the other hand, both groups with DLD produced shorter utterances as measured with MLU_w than their same-age peers with TD (see Table 2). Group differences were highly significant with large effect sizes although substantial variation was seen within all four groups (see Table 2). The O-DLD did not differ significantly from the Y-TD group on MLU_w ($p = 0.211$).

Each TD participant produced two or more finite subordinate clauses and the accuracy of their subordinate clauses was at least 50 percent (see Table 2). Two children from the Y-DLD and two from the O-DLD group did not produce any finite subordinate clauses, but comparisons of DLD and TD groups did not reveal any significant differences with respect to number or accuracy of finite subordinate clauses (see Table 2).

Table 2 displays substantial variation on morphosyntactic accuracy (PGU) in both groups with DLD. However, the PGU range of the Y-DLD did not overlap with the range of the Y-TD group and both groups with DLD had PGU levels markedly below their same-age peers, with large effect sizes (see Table 2). Furthermore, the PGU of the O-DLD group was marginally lower, with a moderate effect size, than the PGU of the much younger Y-TD group ($z = 1.96$, $p = 0.051$, $r = 0.45$).

TABLE 3: MEANS (*M*), STANDARD DEVIATIONS (*SD*) AND RANGE FOR THE ERROR CATEGORIES FOR THE YOUNGER AND OLDER GROUPS WITH DLD OR TD

	Younger DLD (Y-DLD) (*n* = 10)			Younger TD (Y-TD) (*n* = 10)			
	M	***SD***	**Range**	***M***	***SD***	**Range**	
Ungrammatical verb inflection	3.4	3.0	0-10	0.4	0.5	0-1	Y-DLD > Y-TD: z = -2.93, p = 0.004, r = 0.65
Fragment (omission of verb)	1.7	2.8	0-8	0.1	0.3	0-1	Y-DLD ≈ Y-TD: z = -1.96, p = 0.087, r = 0.38
Argument omission	3.3	3.9	0-12	0.3	0.5	0-1	Y-DLD > Y-TD: z = -2.47, p = 0.013, r = 0.55
Word order error	3.3	4.4	0-14	0.2	0.4	0-1	Y-DLD > Y-TD: z = -2.85, p = 0.006, r = 0.64
	Older DLD (O-DLD) (*n* = 9)			**Older TD (O-TD) (*n* =9)**			
	M	***SD***	**Range**	***M***	***SD***	**Range**	
Ungrammatical verb inflection	1.4	1.7	0-5	0	0	0	O-DLD > O-TD: z = -2.85, p = 0.009, r = 0.67
Fragment (omission of verb)	0.9	1.5	0-4	0	0	0	O-DLD ≈ O-TD: p = 0.206
Argument omission	0.3	0.5	0-1	0.3	0.7	0-2	O-DLD ≈ O-TD: p = 1.0
Word order error	2.3	3.2	0-10	0.8	0.8	0-2	O-DLD ≈ O-TD: p = 0.332

3.3 Morphosyntactic errors in the narratives

Besides calculation of the PGU, four specific morphosyntactic error categories were calculated. All but one participant with DLD produced errors of at least one of these types, whereas three participants from the Y-TD and two from the O-TD group did not produce any of these errors. Table 3 displays the very small number of errors of these four types in the TD groups and the large numbers produced by some children with DLD. The Y-DLD group was significantly different from the Y-TD peers with respect to 'ungrammatical verb inflection', 'argument omission' and 'word order errors', and the two groups were marginally different with respect to verb omissions, 'fragments', too (see Table 3). The O-DLD differed significantly from the O-TD group on the measure of ungrammatical verb inflection only. Finally, the O-DLD and Y-TD groups differed significantly on one error type, 'word order errors' ($z = 2.28$, $p = 0.027$, $r = 0.52$).

3.4 Diagnostic accuracy of PGU

Due to very clear group differences, it was relevant to examine the diagnostic accuracy of PGU with ROC-analyses. Table 4 presents the overall diagnostic accuracy ('Area under the curve' calculated in SPSS), optimal cut-offs, and sensitivity and specificity values for the younger and older groups separately (with 95 % confidence intervals calculated with the EBM toolbox at https://ebm-tools.knowledgetranslation.net/calculator/diagnostic/). The diagnostic accuracy was perfect (100 % sensitivity and specificity) for the younger groups; the ten Y-DLD and the ten Y-TD participants were accurately identified as children with and without DLD respectively. The diagnostic accuracy for the older groups was well within the acceptable range (88.9 % sensitivity and specificity); eight of nine participants from the O-DLD and O-TD groups were accurately identified.

TABLE 4: DIAGNOSTIC METRICS FOR PGU AMONG YOUNGER AND OLDER PARTICIPANTS

Area Under the Curve [95 % CI]	Optimal cut-off	Sensitivity (correct DLD) [95 % CI]	Specificity (correct TD) [95 % CI]
Younger children (Y-DLD and Y-TD)			
1.0 [1.0]	74.1	100 % (10/10) [72.2-100]	100 % (10/10) [72.2-100]
Older children (O-DLD and O-TD)			
0.889 [0.726-1.0]	86.9	88.9 % (8/9) [56.5-98.0]	88.9 % (8/9) [56.5-98.0]

4. Discussion

The narrative data from the 38 participants largely provided affirmative answers to the research questions. Even with a small number of participants – resulting in low statistical power – group comparisons revealed differences in morphosyntactic skills between the participants with DLD and their same-age peers with TD. In some cases the older participants with DLD even differed from their younger TD peers.

4.1 Mean length of utterance and use of subordinate clauses by the participants with DLD

In accordance with a number of previous studies (e.g. Duinmeijer et al. 2012) participants with DLD, both Y-DLD and O-DLD, had shorter MLU_w than their same-age peers with TD. MLU_w may be conceptualized as a global measure of expressive language skills (DeThorne et al. 2005). In the light of their need for language intervention and low scores in the background language tests, the shorter MLU_w of the DLD groups is not surprising.

With respect to the use of finite subordinate clauses, statistical comparisons did not reveal any group differences. The reason may be lack of power due to the small group sizes since both groups with DLD had lower accuracy levels and some did not use finite subordinate clauses at all (see Table 2). However, Frizelle and colleagues speculated that children aged 7 to 12 may not use their full linguistic competence if the task does not specifically encourage them to do so. This could also be the case in the present study and that may have affected the more linguistically able participants with TD the most.

4.2 Lower levels of morphosyntactic accuracy in participants with DLD

PGU and specific error types were used for characterizing the morphosyntactic skills of the participants with DLD. The results clearly point to morphosyntax as an area of difficulty and to the narrative generation task as a means of revealing these difficulties.

Lower PGU levels were found for the Y-DLD as well as the O-DLD group relative to the TD controls. PGU seems to tap into core difficulties of children with DLD (Eisenberg & Guo 2013; Fey et al. 2004). Even in the narrative context where participants were relatively free to formulate utterances and choose which items and constructions to include – or avoid – the participants with DLD were less accurate than their TD peers. This is probably due to the high demands for simultaneous processing in narrative generation tasks, which may strain the skills of children with DLD in particular (e.g. Duinmeijer et al. 2012). This results in inaccuracies, even among children who may manage everyday conversations quite well.

The results for the four error types were mixed. 'Ungrammatical verb inflection' – often use of the infinitive or the participle for past tense – was the error type with the clearest differences between the Y-DLD and O-DLD group and their same-age peers. The study thus extends the

context of past tense difficulties from sentence completion and sentence repetition contexts (Vang Christensen & Hansson 2012) to narrative discourse. 'Fragments' (verb omissions) were rare in all participant groups. With the exception of a few individuals with DLD, these school-aged children with DLD complied with the demand for a predicate, a verb, in the utterance, even though they did not always produce the verb in a finite form and even though the Y-DLD group omitted more obligatory arguments than their same-age peers. Finally, in accordance with previous research results from other contexts (Vang Christensen 2019a and Hansson et al. 2003 for Swedish) 'word order errors' were quite common among the participants with DLD. Both the Y-DLD and O-DLD group produced more errors of this type than the Y-TD group.

4.3 PGU as a possible clinical marker of DLD in Danish school-aged children

The PGU measure has shown good potential as a clinical marker task in previous studies of English-speaking children (Eisenberg & Guo 2013; Guo & Schneider 2016). The PGU measure also worked well with the older, Danish-speaking participants in the present study. The large majority of the participants (36 of 38) were classified accurately as DLD or TD by their PGU score. Due to the small number of participants, the confidence intervals are broad (see Table 4), which affects the reliability negatively. Although the results are promising, also from a clinical perspective, they need to be replicated.

4.4 Limitations and future directions

The study provides support for the utility of narrative generation for assessing structural language skills in Danish DLD well into the school years. This is highly relevant for both clinical and research purposes since there is a lack of good standardized tests of morphosyntactic skills in Danish for this age group (see however Vang Christensen 2019a). The results also point to the potential of PGU for identification of DLD in school-aged children. However, all participants with DLD received language intervention at the time of the study, suggesting that their language difficulties were complex or severe; they may not be representative of school-aged children with DLD in the Danish context. As mentioned, further investigation into the utility of the PGU measure for different age groups and for the whole severity range in Danish DLD are needed.

Narratives offer a wealth of data and this study has presented a set of analyses focused on morphosyntax. It seems relevant to characterize the C-units used by the participant groups further, e.g. clausal density or characteristics of verb and noun phrases to gain further insight into characteristics of DLD in Danish and into possible reasons for shorter MLU_w of the DLD compared to TD groups.

Finally, there can be trade-offs between form and content in the narratives of children with DLD. Colozzo, Gillam, Wood, Schnell and Johnston (2011) found that children with DLD were more likely than TD peers to produce either narratives with relatively elaborate content but low morphosyntactic accuracy or narratives with relatively accurate morphosyntax but poor content. Vang Christensen (2019b) investigates, whether this pattern is also found in the data from the participants in the present study.

Acknowledgements

The research behind this paper was undertaken in

1. The project *Language and Cognition – Perspectives from Impairment* (LaCPI) 2011-2014, supported by a grant to Elisabeth Engberg-Pedersen from The Danish Council for Independent Research | Humanities, by The Faculty of Humanities, University of Copenhagen (UCPH), and The Department of Scandinavian Studies and Linguistics (NorS), UCPH. The members of LaCPI also include PhD Hanne Trebbien Daugaard, neuropsychologist Käte From, and eighteen students. I am very grateful for this collaboration.
2. The project *Language Mentor* 2015-2019, supported by a grant to the author by The Health Foundation (Helsefonden) and by NorS, UCPH and involving five SLP students.

Finally, I would like to thank all the children, their parents, the speech-language pathologists and the schools who made the study possible.

References

Bishop, Dorothy V.M. 2010. *Test for Reception of Grammar* (second edition). *Danish Version*. Bromma, Sweden: Pearson Assessment.

Bishop, Dorothy V.M., Margaret J. Snowling, Paul A. Thompson, Trisha Greenhalgh & the CATALISE-2 consortium. 2017. Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry* 58. 1068-1080.

Botting, Nicola. 2002. Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy* 18. 1-21.

Botting, Nicola, Brian Faragher, Zoë Simkin, Emma Knox & Gina Conti-Ramsden. 2001. Predicting pathways of specific language impairment: What differentiates good and poor outcome? *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 42. 1013-1020.

Boudreau, Donna. 2008. Narrative abilities: Advances in research and implications for clinical practice. *Topics in Language Disorders* 28. 99-114.

Colozzo, Paola, Ronald B. Gillam, Megan Wood, Rebecca D. Schnell & Judith Johnston. 2011. Content and form in the narratives of children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 54. 1609-1627.

DeThorne, Laura S., Bonnie W. Johnson & Jane W. Loeb. 2005. A closer look at MLU: What does it really measure? *Clinical Linguistics & Phonetics* 19. 635-648.

Duinmeijer, Iris, Jan de Jong & Annette Scheper. 2012. Narrative abilities, memory and attention in children with a specific language impairment. *International Journal of Language & Communication Disorders* 47. 542-555.

Eisenberg, Sarita L. & Ling-Yu Guo. 2013. Differentiating children with and without language impairment based on grammaticality. *Language, Speech, and Hearing Services in Schools* 44. 20-31.

Eisenberg, Sarita L. & Ling-Yu Guo. 2015. Sample size for measuring grammaticality in preschool children from picture-elicited language samples. *Language, Speech, and Hearing Services in Schools* 46. 81-93.

Elbro, Carsten, Mogens Dalby & Stine Maarbjerg. 2011. Language-learning impairments: a 30-year follow-up of language-impaired children with and without psychiatric, neurological and cognitive difficulties. *International Journal of Language & Communication Disorders* 46. 437-448.

Engberg-Pedersen, Elisabeth & Rikke Vang Christensen. 2017. Mental states and activities in Danish narratives: Children with autism and children with language impairment. *Journal of Child Language* 44. 1192-1217.

Fey, Marc, Hugh W. Catts, Kerry Proctor-Williams, J. Bruce Tomblin & Xuyang Zhang. 2004. Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research* 47. 1301-1318.

Field, Andy. 2009. *Discovering Statistics Using SPSS* (third edition). London: Sage.

Friberg, Jennifer C. 2010. Considerations for test selection: How do validity and reliability impact diagnostic decisions? *Child Language Teaching and Therapy* 26. 77-92.

Frizelle, Pauline, Paul A. Thompson, David McDonald & Dorothy V. M. Bishop. 2018. Growth in syntactic complexity between four years and adulthood: Evidence from a narrative task. *Journal of Child Language* 45. 1174-1197.

Gellert, Anna S & Rikke Vang Christensen. 2012. *Produktivt ordforråd* [Expressive vocabulary]. Unpublished test. Copenhagen: University of Copenhagen.

Guo, Ling-Yu & Phyllis Schneider. (2016). Differentiating school-aged children with and without language impairment using tense and grammaticality measures from a narrative task. *Journal of Speech, Language, and Hearing Research* 59. 317-329.

Hansson, Kristina, Ulrika Nettelbladt & Lawrence B. Leonard. (2000). Specific language impairment in Swedish: The status of verb morphology and word order. *Journal of Speech, Language, and Hearing Research* 43. 848-864.

Hansson, Kristina; Ulrika Nettelbladt & Lawrence B. Leonard. (2003). Indefinite articles and definite forms in Swedish children with specific language impairment. *First Language* 23. 343-362.

IBM Corp. (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.

Jensen de Lopez, Kristine M., Lone Sundahl Olsen & Vasikili Chondrigianni. (2013). Annoying Danish Relatives: Comprehension and production of relative clauses by Danish children with and without SLI. *Journal of Child Language* 41. 51-83.

Johnston. Judith. 2006. *Thinking about Child Language*. Eau Claire, Wisconsin: Thinking Publications.

Leonard, Laurence B. 2014. *Children with Specific Language Impairment* (second edition). Cambridge, MAS: MIT Press.

Lum, Jarrad & Dorthe Bleses. 2012. Declarative and procedural memory in Danish-speaking children with language impairment. *Journal of Communication Disorders* 45. 46-58.

Mayer, Mercer. 1967. *A Boy, a Dog, and a Frog.* New York, NY: Dial Books for Young Readers.

Mayer, Mercer. 1969. *Frog, Where Are You?* New York, NY: Dial Books for Young Readers.

Nippold, Marilyn A., Linda J. Hesketh, Jill K. Duthie & Tracy C. Mansfield. 2005. Conversational versus expository discourse: a study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research* 48. 1048-1064.

Norbury, Courtenay F. & Dorothy V. M. Bishop. 2003. Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders* 38. 287-313.

Norbury, Courtenay F.; Tracey Gemmell; and Rhea Paul. 2014. Pragmatic abilities in narrative production: A cross-disorder comparison. *Journal of Child Language* 41. 485-510.

Norbury, Courtenay F., Debbie Gooch, Charlotte Wray, Gillian Baird, Tony Charman, Emily Simonoff, George Vamvakas & Andrew Pickles. 2016. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry* 57. 1247-1257.

Plante, Elena & Rebecca Vance. 1994. Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools* 25. 15-24.

Semel, Eleanor, Elisabeth H. Wiig & Wayne A. Secord. 2013. *Clinical Evaluation of Language Fundamentals* (fourth edition). Danish Version. Bromma, Sweden: Pearson Assessment.

St. Clair, Michelle C., Andrew Pickles, Kevin Durkin & Gina Conti-Ramsden. 2011. A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI). *Journal of Communication Disorders* 44. 186-199.

Tomblin, J. Bruce, Nancy L. Records, Paula R. Buckwalter, Xuyang Zhang, Elaine Smith & Marlea O'Brien. 1997. The prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research* 40. 1245-1260.

Vang Christensen, Rikke. 2019a. Sentence repetition – a clinical marker for DLD in Danish. Copenhagen: University of Copenhagen, Ms.

Vang Christensen, Rikke. 2019b. Content in relation to form in the narratives of Danish children with developmental language disorder. Copenhagen: University of Copenhagen, Ms.

Vang Christensen, Rikke & Kristina Hansson. 2012. The use and productivity of past tense morphology in specific language impairment: An examination of Danish. *Journal of Speech, Language, and Hearing Research* 55. 1671-1689.

Wechsler, David & Jack A. Naglieri. 2009. *Wechsler Nonverbal Scale of Ability*. Danish Version. Bromma, Sweden: Pearson Assessment.

Wetherell, Danielle, Nicola Botting & Gina Conti-Ramsden. 2007. Narrative in adolescent specific language impairment (SLI): a comparison with peers across two different narrative genres. *International Journal of Language & Communication Disorders* 42. 583-605.