

Grammatical complexity and accuracy of Dutch school-age children with 22q11DS and children with DLD

Student name: Maaike M.T. Steggink
Student number: 3467546
Status: Final
Date: 25 June 2019
Education: Clinical Language, Speech and Hearing Sciences, program in
Clinical Health Sciences, Faculty of Medicine, Utrecht
University, The Netherlands
Supervisors: Dr. T.D. Boerma and Prof. dr. F.N.K. Wijnen
Lecturer: Dr. R. Zwitserlood
Institution: Utrecht Institute of Linguistics OTS, Utrecht, The Netherlands
Wilhelmina Children's Hospital (WKZ) of the University Medical
Center Utrecht, The Netherlands
Style: APA
Intended journal: Journal of Speech, Language and Hearing Research
Reporting guidelines: STROBE¹
Word count: 3798
Word count abstract: 294
Word count Dutch abstract: 284

Abstract

Title: Grammatical complexity and accuracy of Dutch school-age children with 22q11DS and children with DLD.

Background: Almost all children with the 22q11.2 microdeletion syndrome (22q11DS) are delayed in language milestones and language continues to develop slowly. It is evident that these children experience persistent language difficulties, however linguistically informed descriptions of the profile of language impairment are scarce. Moreover, the language delay does not seem directly related to a mental or physical handicap, hearing loss, emotional disorder or environmental deprivation, and therefore some children with 22q11DS receive a diagnosis of DLD.

Aim: In the current study, the grammatical performance of children with 22q11DS was analysed and compared to age-related peers with DLD.

Method: In a cross-sectional, observational pilot study design, 14 children with 22q11DS and 15 children with DLD, aged 6 to 10 years, performed a conversation and picture narrative task. Their elicited language was analysed on six outcome variables, including Mean Length of Utterance in words (MLU-word), Mean Length of Five Longest Utterances (MLU5-word), and four verb-related measures: subject-verb agreement errors, past tense marker errors, dummy auxiliaries and other verb-related errors. Group differences were analysed.

Results: There were no significant differences between the grammatical complexity and accuracy measures of children with 22q11DS and children with DLD.

Conclusion: The current study found no evidence of a difference in performance of children with 22q11DS and children with DLD across grammatical complexity and accuracy measures.

Recommendations: Speech-language therapists should be cautious in diagnosing children with persistent (grammatical) language delays as DLD using a language assessment only, since these could be undiagnosed cases of 22q11DS. It is recommended to refer these children to a multidisciplinary team to evaluate the presence of co-occurring problems, including 22q11DS.

Keywords: 22q11.2DS [MeSH], Developmental Language Disorder [MeSH], language development, speech-language pathology, grammar.

Nederlandse samenvatting

Titel: Grammaticale complexiteit en accuraatheid van Nederlandse schoolgaande kinderen met 22q11DS en kinderen met TOS.

Achtergrond: Bij bijna alle kinderen met het 22q11.2 deletie syndroom (22q11DS) is er sprake van een persisterende taalachterstand. Hoewel het duidelijk is dat deze kinderen taalproblemen ervaren, is er tot op heden weinig bekend over hun taalprofiel. Bovendien lijkt deze taalachterstand niet direct een gevolg van een mentale of fysieke handicap, gehoorverlies, emotionele problemen of een blootstellingsachterstand. Daarom krijgen sommige van deze kinderen ook de diagnose Taalontwikkelingsstoornis (TOS).

Doel: De huidige studie analyseert de grammaticale complexiteit en accuraatheid van schoolgaande kinderen met 22q11DS en vergelijkt hun prestaties met die van leeftijdsgenoten met een TOS.

Methode: In een cross-sectionele, observationele pilot studie met 14 kinderen met 22q11DS en 15 kinderen met TOS, in de leeftijd van 6 tot 10 jaar, is spontane taal ontlokt met behulp van een gesprek en een verteltaak met afbeeldingen. De ontlokte taal is geanalyseerd op zes uitkomstvariabelen: gemiddelde uitingslengte in woorden (MLU-word), gemiddelde uitingslengte van de vijf langste uitingen (MLU5-word), en vier werkwoord-gerelateerde uitkomstmaten: subject-werkwoord agreement fouten, fouten met verleden tijdsformen, dummy werkwoorden en andere werkwoord-gerelateerde fouten. Groepsverschillen zijn geanalyseerd.

Resultaten: De kinderen met 22q11DS verschilden niet significant van de kinderen met TOS op grammaticale complexiteit en accuraatheid.

Conclusie: De huidige studie vond geen bewijs voor een verschil in grammaticale prestaties van kinderen met 22q11DS en kinderen met TOS.

Aanbevelingen: Logopedisten wordt geadviseerd voorzichtig te zijn met het diagnosticeren van een TOS bij kinderen met een hardnekkige (grammaticale) taalachterstand op basis van taalonderzoek alleen. Dit kunnen mogelijk kinderen zijn met 22q11DS. Het is aanbevolen om deze kinderen door te sturen naar een multidisciplinair team om de aanwezigheid van andere stoornissen, waaronder 22q11DS, te laten onderzoeken.

Introduction

Almost all children with the 22q11.2 microdeletion syndrome (22q11DS) are delayed in language milestones and language continues to develop slowly. When comparing their language performance to their intelligence (IQ), language abilities seem to be delayed to a greater extent than one would expect.^{2,3} Therefore, approximately a third to half of the children with 22q11DS receives an additional diagnosis of Developmental Language Disorder (DLD).^{2,4} The question arises how the language performance of children with 22q11DS relates to children with DLD. The current study therefore is the first to systematically analyse grammatical complexity and accuracy in the spontaneous language of school-age children with 22q11DS, and compare their performance to age-related peers with DLD.

22q11DS has also been referred to as DiGeorge syndrome, velocardiofacial syndrome and Shprintzen syndrome. These labels were initially thought to denote different entities, but are now considered variant manifestations of the 22q11.2 microdeletion.² The reported prevalence of 22q11DS varies from 1 in 2000 to 1 in 6000 births. However, it is suspected that 22q11DS remains underdiagnosed in individuals with mild symptoms and the prevalence may approximate 1 in 1000 births, similar to Down's syndrome.^{2,5-7} The clinical phenotype is heterogeneous and may include cardiac defects, palatal abnormalities (cleft palate or velopharyngeal insufficiency), immune deficiencies, feeding problems, mild facial dysmorphism, learning disabilities and speech and language delays. Some children show the first signals of 22q11DS at birth, often as cardiac, feeding or immunological abnormalities. Children with milder symptoms have received the diagnosis when they were looking for help from healthcare professionals for learning disorders, feeding problems or a delayed language and speech development.² The syndrome is diagnosed using the fluorescence in situ hybridization technique (FISH) and children receive the diagnosis at a median age of six and a half years.^{2,6}

Approximately 69% of children with 22q11DS are not yet speaking at the age of 2 and language continues to develop slowly.² Prior studies have found delays in phonology, morphosyntax, receptive vocabulary and narrative abilities in children with 22q11DS.^{2,8,9} These studies however are heterogeneous in participant characteristics (including age, intelligence and control group), outcome variables, study design and sample size, so no clear-cut language profile seems to emerge. Moreover, expressive language is delayed to a greater extent than one would expect on the basis of their cognitive abilities.^{2(p.195)} Furthermore, compared to children with palatal problems of non-syndromic origin, children with 22q11DS perform weaker on receptive language, expressive language, speech-sound inventories and early vocabulary.¹⁰ These findings point to a persistent language disorder that cannot be (fully) explained by cognitive, anatomical or functional (oro-motor) features of

the syndrome, and therefore a third to half of the children with 22q11DS are diagnosed with DLD.

DLD is a language disorder that is not directly related to a mental or physical handicap, hearing loss, emotional disorder or limited exposure to language.¹¹ Hence, a language disorder without an obvious cause. It is characterized by persistent language difficulties that won't resolve spontaneously and interfere with a child's communication and/or educational performance.¹¹ The prevalence of DLD is approximately 7% of kindergarten children.¹² There is general consensus that genetic risk factors play a role in the aetiology of DLD.^{4,5} Because the diagnostic criteria of DLD are exclusionary, rather than inclusionary, the population of children with DLD is heterogeneous.

DLD can affect all language domains including receptive and expressive semantics, pragmatics and phonology, but grammar, especially verb-related grammar, is considered a hallmark deficit.¹⁴ An overview of Dutch studies concerning verb grammar in school-age children with DLD is provided below. De Jong (1999) found that children with DLD of a mean age of 7;8 years selected ditransitive verb frames less often than their typically developing (TD) peers.¹⁵ Secondly, these children used dummy auxiliaries and infinitives as a strategy to circumvent movement and inflection of the main verb. Next, subject-verb agreement was violated more often, by using a singular subject with a plural verb, bare stem form or root infinitive. In past tense context children with DLD would often omit the tense marker or use the present tense form instead. In another study by Wexler et al. (2004), children with DLD (aged 6;0 – 8;2) produced more root infinitives than younger TD children (aged 3;1 – 3;7).¹⁶ Verhoeven et al. (2011) found omission of the third person singular agreement marker to be a clinical marker of DLD.¹⁷

So far, one study has compared language characteristics of children with 22q11DS to children with DLD.¹⁸ Kambaros & Grohmann (2017) performed a case study of a boy with 22q11DS. They compared him to seven peers with DLD at 6 years of age, and to nine peers with DLD at 10 years of age. The boy with 22q11DS could not be differentiated from the children with DLD on global, receptive and expressive language test scores (quotients) at both time points. However, at the age of 6, his sentence length (expressed as MLU-word, Mean Length of Utterances in words) was higher than the mean MLU-word of his peers with DLD. Furthermore, also at the age of 6, he scored worse than his peers with DLD on sentence repetition of subject relative clauses. This study suggests that the language profile of 22q11DS could be similar to that of children with DLD, however, since it is a case study, the level of evidence to support this claim is low.

The current study aims to investigate whether children with 22q11DS present a similar grammatical profile to that of children with DLD. Linguistically informed descriptions of children with 22q11DS can help speech-language therapists (SLTs) in their assessment and

therapy work. A language profile could aid in identifying cases of 22q11DS in the DLD population that have not yet been diagnosed. Moreover, it can guide towards appropriate treatment, more specifically, to evaluate if grammatical therapy programmes designed for children with DLD would be fitting for children with 22q11DS. The research question is as follows:

What are the similarities and differences in grammatical complexity and accuracy (including MLU-word, MLU5-word, root infinitives, dummy auxiliaries, ditransitive verbs, subject-verb agreement errors, past tense marker errors and other verb-related errors) of school-age children with 22q11DS, compared to age-related peers with DLD?

As mentioned earlier, no clear-cut hypotheses concerning grammatical development in children with 22q11DS seem to emerge from prior work. A cross-sectional study by Persson et al. found a shorter average utterance length for 6- to 10-year-old children with 22q11DS compared to TD children.⁹ The 22q11DS-DLD comparison study of Kambanaros et al. found that the boy with 22q11DS had a shorter MLU-word at the age of 6, but not at the age of 10, compared to children with DLD.¹⁸ No other studies have addressed grammatical development of children with 22q11DS in relation to children with DLD.

Method

Participants

Children with 22q11DS were recruited via the Wilhelmina Children's hospital (WKZ)/University Medical Center Utrecht (UMCU), and the national parental support group Stichting Steun 22q11 in The Netherlands. The children with DLD were recruited through the Royal Auris Group, an organisation that provides care and education services for children with communication problems. Participants were selected based on the following inclusion criteria: (1) chronological age of 6-10 years; (2) 22q11DS group: a diagnosis of 22q11DS confirmed by FISH.⁵ (3) DLD group: a diagnosis of DLD in agreement with the *Siméa richtlijn voor toelaatbaarheid*.¹⁹ The following exclusion criteria were used: (1) hearing loss of >35 dB at the time of testing; (2) performal and verbal IQ scores of <70, to control for language delays related to (significant) intellectual impairment.²⁰ (3) DLD group: other comorbidities that affect language development, including neurological damage, autism or ADHD.²⁰ The flowchart in figure 1 indicates the amount of eligible participants and the selection.

[Figure 1]

Materials and procedures

Children were assessed from January to August 2018 at the UMCU by a team of researchers with a degree in special education and/or linguistics. Ethical approval was obtained from the Medical Research Ethics review board of the UMCU (approval number: NL62366.041.17), as part of a larger study that aims to compare the language profile of children with 22q11DS to children with DLD and TD children. The participating children and their parents gave informed consent right before the language assessment.

Spontaneous language assessment. Spontaneous language was elicited in a conversation and in a narrative task. A member of the research team first interviewed the child about decontextualized topics close to the interests of the child (e.g. summer holidays, pets, birthdays, favourite games) to elicit spontaneous language about things they had seen or experienced. Next, a picture narrative was elicited using the Dutch version of the Multilingual Assessment Instrument of Narratives (MAIN).²¹ The investigator modelled the task with the picture sequence *Cat*. Then, the picture sequence *Baby birds* was presented to the child. The child was instructed to look at the pictures and tell the story to the investigator, who could not see the pictures. The investigator did not ask any questions but was allowed to encourage the child.

Spontaneous language analysis. The audio fragments of the dialogue and MAIN narrative were transcribed and coded by the first author. Beforehand, a transcription protocol was formulated (see appendix A), consistent with the Codes for the Human Analysis of Transcripts (CHAT) transcription system.²² An utterance was defined as a T-unit: a main clause and all subordinate clauses attached to it.²³ The researchers did not know if a child had DLD or 22q11DS during the transcription process.

10% of the annotations, consisting of a total of 473 T-units, were checked by a second researcher. The transcripts were randomly selected from the two participant groups and originated from three children with 22q11DS and three children with DLD. Disagreements were discussed and a checklist was formulated (see Appendix B) to increase annotation accuracy. The first transcriber then corrected all transcriptions following the guidelines in this checklist. After this process, annotation consensus reached 91,5% of T-units on the checked transcripts.

Background variables. IQ was assessed with either a verbal or nonverbal IQ test. The physician, hospital, school or institution that was involved with the child was asked to provide the results of the intelligence assessment. If intelligence was not assessed before or results were outdated, the research team applied a standardized IQ test.

The Recalling Sentences (Zinnen Herhalen) subtest of the Clinical Evaluation of Language Fundamentals-4-NL (CELF-4-NL: ZH) was administered to validate that all participants have a language disorder.²⁴ Sentence repetition tasks are considered a reliable

clinical marker of a language difficulties.^{25,26} In this task, a child listens to spoken sentences of increasing length and complexity and repeats the sentences without changing their content. The mean of the normed score is 10 with an SD of 3.

The Peabody Picture Vocabulary Task (PPVT), a standardized receptive vocabulary test, was administered as well.²⁷ The child is asked to select out of four pictures the one that reflects the meaning of a word spoken by the examiner. The total score is converted to an quotient score with a mean of 100 and standard deviation (SD) of 15.

Study design

The grammatical complexity and accuracy of children with 22q11DS and children with DLD was compared in a cross-sectional observational design. An independent measures design was used with one independent variable of group (22q11DS and DLD) and eight dependent variables consisting of grammatical complexity and accuracy measures (table 1). MLU-word and MLU5-word were included as general measures of grammatical complexity. Six verb-related grammar outcome variables were included, including subject-verb agreement errors, past tense marker errors, root infinitives, dummy auxiliaries, ditransitive verbs and other verb-related errors. Due to the varying number of utterances elicited from the children, the number of occurrences of each marker was divided by the number of obligatory contexts, resulting in percentage-type variables.

Chronological age at the moment of testing, gender, total IQ, receptive vocabulary and sentence repetition were included as participant variables.²⁸⁻³¹

[Table 1]

Analysis

Data from the transcripts were analysed using Computerized Language Analysis Software (CLAN).³² The commands used in this study are provided in Appendix C.

Data was analysed statistically with IBM SPSS Statistics 25.³³ A MANOVA was used to compare aspects of grammar (see table 1) of the 22q11DS group to the DLD group, using an alpha of .05. Pillai's trace was chosen as the test statistic. The assumptions of normality and heterogeneity of variance were checked with Kolmogorov-Smirnov's test and Levene's test of Equality of Error Variances, respectively. If the MANOVA's results were significant, follow-up analyses of univariate ANOVAs with post-hoc Bonferroni corrections were applied.

For variables that violated assumptions for a MANOVA, nonparametric Mann-Whitney U tests were applied. If more than one variable would be non-normal, post-hoc Bonferroni corrections were applied to correct for multiple testing.

Effect sizes are provided using Pearson's correlation coefficient r , with $r = .10$ constituting a small effect, $r = .30$ a medium effect and $r = .50$ a large effect.³⁴

Results

Participant characteristics

Twenty-nine 6 to 10 year old Dutch school-age children were included, of which 14 with a diagnosis of 22q11DS and 15 with DLD (see table 2). All children passed the hearing screening. For one child in the DLD group, total IQ was missing. For one child in the 22q11DS group, receptive vocabulary was missing.

Children with 22q11DS had a mean chronological age of 8 years and 6 months. For the children with DLD this was 8 years and 1 month. The groups did not differ significantly in gender, $\chi^2(1) = 0.042, p > .05$, or chronological age, $U = 80, z = -1.092, p > .05$. Group differences of total IQ were not analysed since four different IQ tests were administered across the participants, including both verbal and non-verbal versions.

The DLD group's mean IQ score was 105 and within normal limits. The children with 22q11DS as a group had a below average total IQ of 74. On the Recalling Sentences subtest, both children with DLD and 22q11DS performed below average (3 and 5, respectively). The receptive vocabulary of children with DLD was 93 and within normal limits. The children with 22q11DS had a mean receptive vocabulary of 83. This is considered below average.

[Table 2]

Assumptions

The assumption of normality was met for the variables MLU-word, MLU5-word, subject-verb agreement errors, verb-related errors and dummy auxiliaries. Past tense marker errors, root infinitives and ditransitive verbs were significantly non-normal.

For all dependent variables, the variances were equal for the DLD and 22q11DS group, except for proportion of ditransitive verbs, $F(1, 27) = 5.08, p < .05$. The variance-covariance matrices were equal across groups, $M(28, 2511) = 46,73, p > .05$.

Because past tense marker errors did not meet the assumption of normality, this variable was tested nonparametrically with a Mann-Whitney U test. Root infinitives and ditransitive verbs were summarized descriptively but excluded from the analysis, because of their low occurrence in the transcriptions (17 and 10 annotations, respectively).

Number of T-units, number of words, number of T-units containing a verb

A total of 2537 T-units were analysed, of which 1132 were produced by children with 22q11DS and 1405 by children with DLD. A total of 12211 words were analysed, of which 5538 in the 22q11DS group and 6673 of the DLD group. In the 22q11DS group, 60.8% ($SD = 9.9$) of T-units contained a verb phrase. In the DLD group, this percentage was 60.2% ($SD = 16.5$).

Grammatical complexity and accuracy

Using Pillai's trace, there were no significant differences between children with DLD and children with 22q11DS on measures of MLU-word, MLU5-word, proportion of subject-verb agreement errors, proportion of verb-related errors and proportion of dummy auxiliaries, $V = 0.03$, $F(5, 23) = 0.160$, $p > .05$, $r = 0.034$.

[Table 4]

MLU-word and MLU5-word. Children with DLD had a slightly lower mean MLU of 5.25 words, compared to the children with 22q11DS, who had a mean MLU of 5.44 words. However, this was not significant. Their five longest T-units had a mean length of 12.57 words and 12.12 words, respectively.

Subject-verb agreement errors. Children with 22q11DS violated subject-verb agreement in 5.9% of the T-units containing a subject and a verb. For children with DLD, the error rate was 5.8%.

Other verb-related errors. Children with 22q11DS made 17.5% verb-related errors in T-units containing a verb phrase. Children with DLD made 20.1% verb-related errors.

Dummy auxiliaries. Children with 22q11DS used dummy auxiliaries in 6.9% of all T-units containing a verb phrase; children with DLD used dummy auxiliaries in 7.4% of the verb phrases.

Past tense marker errors

The proportion of past tense marker errors did not differ significantly between children with DLD ($M = 0.185$, $Mdn = 2.0$) and children with 22q11DS ($M = 0.168$, $Mdn = 0.1$), $U = 100$, $z = -0.226$, ns , $r = 5.39$. Children with 22q11DS did not use a past tense form in 18.5% of the contexts where it was expected. For children with DLD, this rate was 16.9%. Individual variation between participants was high, 95% CI [1.6, 32.2] for children with 22q11DS; 95% CI [2.3, 34.7] for children with DLD.

Root infinitives and ditransitive verbs

For children with 22q11DS, root infinitives occurred in 0.8% of the T-units containing a verb phrase. For children with DLD, this rate was 1.7%.

Ditransitive verbs occurred in 0.3% of the T-units contained a verb in the 22q11DS group and in 1.2% of the T-units with a verb of the DLD group.

Discussion

Children with 22q11DS experience delayed and persistent language difficulties and so far linguistically informed descriptions of their language profile are scarce. Therefore, the current study aimed to describe the grammatical language profile of school-age children with 22q11DS. Since their language delay does not seem directly related to a mental or physical handicap, hearing loss, emotional disorder or environmental deprivation, and therefore children with 22q11DS often receive a diagnosis of DLD, their performance was compared to age-related peers with DLD. The current study looked into grammar, because this is considered to be a hallmark deficit of children with DLD.

The current study found no evidence of a difference in performance of children with 22q11DS and children with DLD across verb grammar measures, including MLU-word, MLU5-word, dummy auxiliaries, past tense marker errors and other verb-related errors (subject drop, object drop, errors in word order and overregularization of verb inflection). Although the aetiology of 22q11DS and DLD is different, the grammatical profiles seem to overlap. This is consistent with the case study of Kambaranos et al. (2017), who did not find a difference in productive morphosyntax of a boy with 22q11DS and peers with DLD at 10 years of age.¹⁸

Since the current study lacked a control group of TD children, we can only speculate if the grammatical profile of children with 22q11DS deviates from the norm. However, we do suspect this is the case. The children from the current study had a below average score on the Recalling Sentences subtest of the CELF-4-NL, which indicates that language is indeed disordered.²⁶ Furthermore, children with 22q11DS have notable weaker grammar scores compared to Dutch TD children of a similar age in previous studies. Zwitserlood et al. (2014) found a mean MLU-word of 6.84 for TD children at the age of 8;6 years (MLU-word of our 22q11DS group was 5.44). Also, his verb-related error category (which was defined more broadly than ours, and included subject-verb agreement errors, past tense agreement errors and root infinitives too), found an error rate for TD children of 3.9%, which is remarkably lower than the 17,5% of our 22q11DS group.³⁵ De Jong found a 99% correct score for past tense marking in TD children of 7;6 years of age (our children with 22q11DS had an error

rate of 18.5%).¹⁵ These data suggest that children with 22q11DS have a lower grammatical complexity and accuracy compared to TD children of a similar age.

A notable outcome concerns IQ. The grammatical profiles of the children are similar and yet, the children with DLD as a group had a mean IQ 31 points higher than the children with 22q11DS. Considering the 22q11DS group, their grammatical performance could be a result of their below average IQ. After all, speaking in longer and more complex sentences places a higher demand on cognitive resources, e.g. working memory.³⁶ However, for children with DLD, the aetiology is different since their IQ is within normal limits but their grammatical performance is below average. It is still unclear how cognitive mechanisms influence and shape language development. Furthermore, if grammar is vulnerable in both DLD and 22q11DS, the question arises how children with other types of language disorders will perform. Grammar might be a shared problem area across types of language disorders.

Current study aimed to profile the grammar of school-age children with 22q11DS. However, to be able formulate a comprehensive language profile for these children, it is recommended that future studies investigate other language domains and other age groups. Moreover, since children with 22q11DS seem to have a grammatical profile similar to children with DLD, future studies should evaluate if grammatical therapy designed for children with DLD could be beneficiary for children with 22q11DS too.

Strengths and limitations

The current study applied a linguistic approach to study language development in children with 22q11DS. It is the first to use spontaneous language analysis to study grammar in children with 22q11DS. This gives a more detailed analysis of grammatical development than a score on a standardized language test. Furthermore, it is the second comparative language study on 22q11DS and DLD and the first to include a group of children with 22q11DS.

There were a few limitations to the current study and the results should therefore be treated with some caution. Firstly, the sample size is relatively small and no sample size calculation was done prior to data collection. The fact that we did not find differences between the two groups of children could be due to insufficient statistical power. Secondly, due to the timeframe in which the study had to be completed, no control group of TD children was included.

Clinical implications

The current study emphasizes that school-age children with 22q11DS experience language delays and that they cannot be differentiated from their peers with DLD on measures of grammar. SLTs therefore should be cautious in diagnosing DLD using a language assessment only and should consult a multidisciplinary team to evaluate the presence of co-

occurring problems, including 22q11DS. This practice will aid in identifying children with mild cases of 22q11DS that are at risk of being underdiagnosed. Identifying 22q11DS in a child is essential to monitor and treat the complex medical and behavioural problems present in this population.

Conclusions

The current study found no evidence that school-age children with 22q11DS perform differently from children with DLD on grammatical complexity and accuracy in their spontaneous language, including MLU-word, MLU5-word, dummy auxiliaries, past tense marker errors and other verb-related errors (subject drop, object drop, errors in word order and overregularization of verb inflection). Although the aetiology of 22q11DS and DLD is different, grammatical profiles are similar.

References

1. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007;147(8):573–7.
2. Solot CB, Knightly C, Handler SD, Gerdes M, McDonald-McGinn DM, Moss E, et al. Communication Disorders in the 22q11.2 Microdeletion Syndrome. *J Commun Disord.* 2000;33(215):187–204.
3. Glaser B, Mumme DL, Blasey C, Morris MA, Dahoun SP, Antonarakis SE, et al. Language skills in children with velocardiofacial syndrome (deletion 22q11.2). *J Pediatr.* 2002;140(6):753–8.
4. Goorhuis-Brouwer SM, Dikkers FG, Robinson PH, Kerstjens-Frederikse WS. Specific language impairment in children with velocardiofacial syndrome: four case studies. *Cleft palate-craniofacial J.* 2003;40(2):190–5.
5. K.J. Golding-Kushner. Speech and language disorders in velo-cardio-facial syndrome. In: Murphy KC, Scambler PJ, editors. *Velo-Cardio-Facial Syndrome: A Model for Understanding Microdeletion Disorders.* 1st ed. Cambridge University Press; 2005.
6. Botto LD, May K, Fernhoff PM, Correa A, Coleman K, Rasmussen SA, et al. A Population-Based Study of the 22q11.2 Deletion: Phenotype, Incidence, and Contribution to Major Birth Defects in the Population. *Pediatrics.* 2003 Jul 1;112(1):101–7.
7. Niklasson L, Rasmussen P, Oskarsdóttir S, Gillberg C. Neuropsychiatric disorders in the 22q11 deletion syndrome. *Genet Med.* 2001;3(1):79–84.
8. Van den Heuvel E. Pragmatic language skills in children with 22q11.2 deletion syndrome and Williams syndrome: A cross-sectional and prospective longitudinal study. KU Leuven; 2016.
9. Persson C, Niklasson L, Óskarsdóttir S, Johansson S, Jönsson R, Söderpalm E. Language skills in 5-8-year-old children with 22q11 deletion syndrome. *Int J Lang Commun Disord.* 2006;41(3):313–33.
10. Scherer NJ, D'Antonio LL, Kalbfleisch JH. Early speech and language development in children with velocardiofacial syndrome. *Am J Med Genet.* 1999;88(6):714–23.
11. Bishop DVM. Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *Int J Lang Commun Disord.* 2017;52(6):671–80.
12. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *J speech, Lang Hear Res.*

- 1997;40(6):1245–60.
13. Bishop DVM. Genetic and environmental risks for specific language impairment in children. *Philos Trans R Soc London B Biol Sci.* 2001;356(1407):369–80.
 14. Conti-Ramsden G, Jones M. Verb use in specific language impairment. *J Speech, Lang Hear Res.* 1997;40(6):1298–313.
 15. De Jong J. Specific language impairment in Dutch: Inflectional morphology and argument structure. University of Groningen; 1999.
 16. Wexler K, Schaeffer J, Bol G. Verbal syntax and morphology in typically developing Dutch children and children with SLI: How developmental data can play an important role in morphological theory. *Syntax.* 2004;7(2):148–98.
 17. Verhoeven L, Steenge J, van Balkom H. Verb morphology as clinical marker of specific language impairment: Evidence from first and second language learners. *Res Dev Disabil.* 2011;32(3):1186–93.
 18. Kambaranos M, Grohmann KK. Linguistic and Nonverbal Abilities over Time in a Child Case of 22q11 Deletion Syndrome. *Biolinguistics.* 2017;11:57–82.
 19. Stichting Siméa. Siméa richtlijn toelaatbaarheid [Internet]. 2017 [cited 2018 Nov 30]. Available from: www.simea.nl
 20. Nederlandse Vereniging voor Logopedie en Foniatrie. Richtlijn Logopedie bij taalontwikkelingsstoornissen [Internet]. Woerden; 2017 [cited 2018 Nov 30]. Available from: <https://www.nvlf.nl/paginas/openbaar/vakgebied/vakinhoud/richtlijn-tos>
 21. Gagarina N, Klop D, Kunnari S, Tantele K, Välimaa T, Balciuniene I. Multilingual Assessment Instrument for Narratives (MAIN). *ZAS papers in linguistics* 56. Berlin: ZAS. Retrieved from <http://www.zas.gwz-berlin.de/zaspil56.html>; 2012.
 22. MacWhinney B. The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs. Psychology Press; 2014.
 23. Hunt KW. Grammatical Structures Written at Three Grade Levels. *NCTE Research Report No. 3.* 1965;
 24. Kort W, Schittekatte M, Compaan E. CELF-4-NL: Clinical Evaluation of Language Fundamentals. Amsterdam Pearson Assess Inf. 2008;
 25. Conti-Ramsden G, Botting N, Faragher B. Psycholinguistic markers for specific language impairment (SLI). *J child Psychol psychiatry.* 2001;42(6):741–8.
 26. Klem M, Melby-Lervåg M, Hagtvet B, Lyster S-AH, Gustafsson J-E, Hulme C. Sentence repetition is a measure of children's language skills rather than working memory limitations. *Dev Sci.* 2014/07/01. 2015 Jan;18(1):146–54.
 27. Schlichting L. Peabody Picture Vocabulary Test, Dutch version. Amsterdam: Pearson Assessment and Information BV; 2005.
 28. Hendriksen JGM, Hurks PPM. WPPSI-III-NL Wechsler preschool and primary scale of

- intelligence; nederlandse bewerking. 2009;
- 29. Kort W, Schittekatte M, Dekker PH, Verhaeghe P, Compaan EL, Bosmans M, et al. WISC-III NL Wechsler intelligence scale for children. Derde Editie NL. Handleiding en Verantwoording. Amsterdam Psychol HTPNIV. 2005;
 - 30. Snijders JT, Snijders-Oomen AWM, Tellegen PJ. SON-R 21/2-7: Snijders-Oomen niet-verbale intelligentietest: handleiding en verantwoording. Swets Test Publishers; 1998.
 - 31. Wechsler D. Naglieri. JA (2008) WNV NL. Wechsler Nonverbal Scale of ability. Nederlandstalige bewerking. Amsterdam: Pearson;
 - 32. MacWhinney B, Wagner J. Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprachsforsch Online-Zeitschrift zur verbalen Interaktion*. 2010;11:154.
 - 33. IBM Corp. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp; 2017.
 - 34. Field A. *Discovering Statistics Using SPSS*. SAGE Publications; 2009.
 - 35. Zwitserlood R, van Weerdenburg M, Verhoeven L, Wijnen F. Development of morphosyntactic accuracy and grammatical complexity in Dutch school-age children with SLI. *J Speech, Lang Hear Res*. 2015;58(3):891–905.
 - 36. Deák GO. Interrelations of language and cognitive development. *Encycl Lang Dev*. 2014;284–91.

Tables and figures

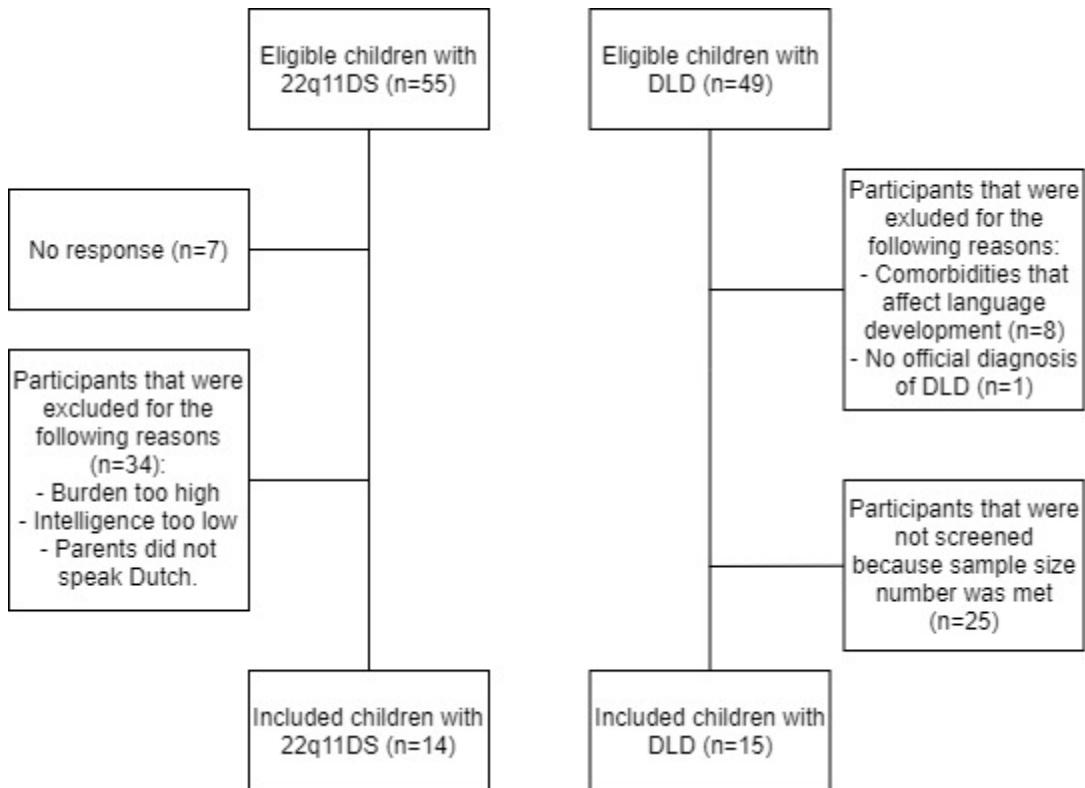


Figure 1. Flowchart of study participation.

Table 1

Study parameters.

Variables	Operationalization
MLU-word	Mean number of words per T-unit.
MLU5-word	Mean number of words of the five longest T-units.
% Subject-verb agreement errors	Number of subject-verb agreement errors, divided by sum of T-units containing a subject and a verb. The subject of a sentence is the actor or idea of the sentence. The verb is the action. Subjects and verbs need to agree in number, i.e. both should be singular or both should be plural. <i>*jullie loop (*you walk).</i> <i>*hij vangen de kat (*he catch the cat)</i>
% Dummy auxiliaries	Number of dummy auxiliaries, divided by the sum of T-units containing a verb phrase. Sentences with dummy auxiliaries contain a semantically empty auxiliary (often go/gaan or do/doet) and an infinitive. Children use dummy auxiliaries to avoid inflection of the main verb. <i>hij doet fietsen (he does cycling).</i> <i>kat gaat slapen (cat goes sleeping)</i> <i>zij is aan het wassen (she is washing)</i>
% Past tense marker errors	Number of T-units with an incorrectly tensed verb in the context of a past tense adverb (such as <i>yesterday</i> (gisteren) or <i>when</i> (toen)) or sentence context (i.e. a description of a past event), divided by the sum of t-units in past tense. <i>*gisteren valt mama (*yesterday fall mama).</i>
% Other verb-related errors	Number of other verb-related errors, including subject, object or auxiliary omissions, errors in word order and overregularization of verb inflection, divided by the sum of T-units containing a verb phrase. Cases of topicalization and legitimate/adult-like object/subject drop were excluded.
% Root infinitives	Number of root infinitives, divided by the sum of T-units containing a verb phrase. A root infinitive is an infinite verb, i.e. a verb without inflection. <i>*hij het koekje eten (*he the cookie eat).</i>
% Ditransitive verbs	Number of ditransitive verbs, divided by the sum of T-units containing a verb phrase. Ditransitive verbs are verbs that takes a subject and two objects. <u>Jan geeft een kado aan Marie</u> (<i>John gives a present to Mary</i>).

Table 2

Participant characteristics.

Characteristics	22q11DS (n=14)			DLD (n=15)		
	M	SD	Range (min, max)	M	SD	Range (min, max)
Age (in months)	102.15	18.16	81 - 132	97.27	20.73	74 - 132
Gender (number of boys)	8			8		
Total IQ	74.00	8.60	64 - 94	105.43	15.70	85 - 136
WPPSI-III-NL ^a	67.50	0.71				
WISC-III-NL ^b	73.40	7.25		113.50	19.09	
SON-R ^c	83.50	14.85		102.00	7.17	
WNV-NL ^d				105.13	18.93	
Receptive Vocabulary (PPVT)	83.08	13.70	66 - 110	93.20	13.56	72 - 117
Recalling Sentences (CELF-4-NL: ZH)	5.15	2.27	1 - 8	3.93	1.98	1 - 7

Note. M = mean; SD = standard deviation; WPPSI-III = Wechsler Preschool and Primary Scale of Intelligence-III-NL; WISC-III-NL = Wechsler Intelligence Scale for Children-III-NL; SON-R = Snijders-Oomen Niet-Verbale Intelligentietest; WNV-NL = Wechsler Non Verbal-NL; PPVT = Peabody Picture Vocabulary Task; CELF-4-NL: ZH = Clinical Evaluation of Language Fundamentals-4-NL: Zinnen Herhalen.

^an=2. ^bDLD n=2 and 22q11DS n=10. ^cDLD n=4 and 22q11DS n=2. ^dn=8.

Table 4

Grammatical complexity and accuracy measures as percentages of sums of T-units in children with 22q11DS and children with DLD.

Variables	22q11DS (n=14)			DLD (n=15)		
	M (SD)	95% CI	Range (min, max)	M (SD)	95% CI	Range (min, max)
MLU-word	5.44 (0.94)	4.90 - 5.98	3.95 - 6.86	5.25 (1.25)	4.56 - 5.96	3.47 - 7.21
MLU5-word	12.57 (2.71)	11.00 - 14.14	8.20 - 17.00	12.12 (3.33)	10.28 - 13.96	7.60 - 17.80
% Subject-verb agreement errors ^b	5.9 (4.0)	3.6 - 8.3	0 - 14.0	5.8 (3.0)	4.1 - 7.5	2.00 - 14.0
% Dummy auxiliaries ^c	6.9 (5.1)	3.4 - 10.0	0 - 16.0	7.4 (4.7)	4.8 - 10.0	0 - 15.0
% Past tense marker errors ^d	16.9 (26.5)	1.6 - 32.2	0 - 100.0	18.5 (29.3)	2.3 - 34.7	0 - 100.0
% Other verb-related errors ^c	17.5 (9.6)	12.1 - 22.8	6.00 - 34.0	20.1 (7.3)	16.1 - 24.2	10.00 - 40.0
% Root infinitives ^c	0.8 (1.3)	0.1 - 1.5	0 - 4.0	1.7 (2.7)	0.2 - 3.2	0 - 10.0
% Ditransitive verbs ^c	0.3 (0.9)	-0.1 - 0.8	0 - 3.0	1.2 (1.5)	0.0 - 2.1	0 - 5.0

Note. M = mean; SD = standard deviation; CI = confidence interval.

^acounts/sum of T-units. ^bcounts/sum of T-units containing a subject and a verb. ^ccounts/sum of T-units containing a verb. ^dcounts/sum of T-units with a past tense adverb or in a past tense context.

Appendix A

Transcription protocol (in Dutch)

THESIS MAAIKE GUIDELINES TRANSCRIBEREN (OP BASIS VAN CoDEMBI)

Inleiding

Het Child Language Data Exchange System (CHILDES) bestaat uit twee componenten: De database, waarin je transcripties (in CHAT formaat) vindt van allerlei taalopnames en CLAN, een programma waarmee de CHAT-bestanden geanalyseerd kunnen worden en bij voorbeeld frequenties van woorden of woordcombinaties berekend kunnen worden, of de MLU (mean length of utterance: het gemiddeld aantal woorden per uiting). Als je het programma zelf wilt installeren, let er dan op dat je de voor jouw configuratie geschikte versie van het net plukt.

Het gegevensbestand

Op de CHILDES-thuispagina kun je veel informatie vinden over CHILDES en eerstetaalverwerving in het algemeen: childepsy.cmu.edu

Vanaf de thuispagina kun je de programmatuur downloaden voor CHILDES, je vindt er de handleidingen om te leren werken met CHAT en CLAN, er is een bibliografie om literatuur te zoeken en je hebt toegang tot het gegevensbestand: in het gegevensbestand vind je een grote verzameling transcripties van opnamen in diverse talen, er zijn zowel dwarsdoorsnede- als longitudinale opnamen met meer en minder gestuurde taalproductie. Een verzameling van transcripties die door één onderzoeker of onderzoeksgroep aangeleverd zijn, noemen we een *corpus*.

De CHILDES-editor; CHAT-bestanden

Alle transcripties in CHILDES hebben een vergelijkbare structuur. Alleen transcripties in deze vorm, het zogenaamde CHAT-formaat, kunnen door de analyseprogramma's (CLAN dus) 'gelezen' worden.

Bestanden moeten worden opgeslagen als tekst bestanden en eindigen op .cha.

Een transcriptie bevat uiteraard de uitgetypte spraak, maar hij begint verplicht met de volgende kopjes (obligatory headers):

HEADERS:

@Begin
@Languages: nld
@Participants CHI child, INT interviewer
@Options
@ID: CHI child 1, INT interviewer
@Transcriber: (geef hier je naam aan zodat we weten van wie de transcriptie is)

Iedere deelnemer wordt aangegeven met drie hoofdletters (de speaker ID), daarna de naam en vervolgens de 'rol' van de deelnemer, bijvoorbeeld, child, investigator. De verschillende deelnemers zijn gescheiden door een komma.

Na de achtergrondinformatie volgt dan eindelijk wat er daadwerkelijk gezegd is. Dit worden de sprekersregels genoemd (main speaker tiers). Hiervoor gelden de volgende conventies:

Sprekersregels beginnen altijd met een *, de drie hoofdletters van de deelnemer (Speaker ID), dubbele punt, [tab], dus: *CHI:[tab]uiting.

Iedere uiting begint op een nieuwe regel. Iedere sprekersregel moet eindigen met een punt, een vraagteken of een uitroepstreken en vervolgens een [return].

Gebruik voor de interviewer altijd *INT en voor het kind altijd *CHI

Helemaal aan het eind van het transcript volgt nog verplicht: @End

Dit geeft het einde van de transcriptie aan en is weer noodzakelijk voor de analyseprogramma's.

Geef het bestand als naam het nummer van het kind, je naam en de datum in yyyyymmdd, bv.
1_maaika_20190205.cha.

SHORTLIST GEBRUIKTE CHILDES-NOTATIES IN DE TRANSCRIPTIES

.	einde uiting
?	einde uiting vraagintonatie of vraagconstructie
!	einde imperatieve uiting/gebiedende wijs (geen exclamaties)
+...	onafgemaakte uiting (trailing off, staat aan het einde van een uiting). Zin wordt wel meegenomen in MLU berekening.
+,	uiting door spreker afgemaakt (staat aan het begin van een uiting)
++	uiting door iemand anders dan spreker afgemaakt (staat aan het begin van een uiting)
+//.	zelfinterruptie (staat aan het eind van een uiting)
+/.	interruptie door andere spreker (staat aan het eind van een uiting)
+”.	quote (aan het einde van de voorafgaande uiting)
+”	begin quote (voor quotatie)
&x	fonologisch fragment, bijv. &e (wordt niet meegenomen in MLU berekening)
<tekst tekst>	groep woorden waarop volgende het symbool tussen [] betrekking op heeft
_	gelinkte woorden of chunks (bijv. mickey_mouse)
tekst(tekst)tekst	reductie (niet-afwijkend van target)
[]	repetitie/herhaling.
["]	letterlijke herhaling, bijv. als het kind de woorden van de onderzoeker letterlijk herhaalt
[//]	retracing/herformulering, met correctie
[///]	complete herformulering, zonder correctie
&=actie	voor acties, bijvoorbeeld &=lacht, &=hoest, enz.
eh@fp	filled pauses: eh, ehm, mmm, dus één code voor alle vormen
@i	oh, aha, noh, hm, ja, nee, yeah, nou, he, etc. (hm@i hm@i)
@c	child-invented form, e.g. prikprukprikk@c of spelgeld@c
@g	general special form ((huppekee@g)
@o	onomatopoeias: include animal sounds and attempts to imitate natural sounds (woefwoef@o, bong@ tegen de boom aan, flapper_de_flap@o)
@I	letters, bijvoorbeeld: n@I I@I (NL)
(.)	pauze
(..)	lange pauze
(...)	heel lange pauze
[?]	beste gok
[*]	fout, direct na de fout en met correctie, bijv: wij loop [: lopen] [*] naar huis
xxx	onverstaanbare spraak
[: tekst]	juiste vorm (in de volwassen taal), bijv. hij heb [: heeft] .
<tekst>	(idem voor) alle woorden die tussen de haken staan, bijv. <ik weet> [/] ik weet het niet .
Naam_Kind	in plaats van de naam van het kind, om anonimiteit te waarborgen.

%MAA

codeer het volgende in de %maa tier:

NB: codeer alleen volledige zinnen. Incomplete zinnen markeer je met +... (trailing off), +/- (zelfinterruption) of +/. (interruption door andere spreker). Imperatieveen codeer je ook niet.

VP+ zin met een werkwoord

VP- zin zonder een werkwoord

ELL elliptische zin (vaak antwoord op vragen van de interviewer). Bijv:

*CHI: ja@i .

%maa: VP- ELL

Agreement:

noteer persoon en getal. doe dit per zin (dus in het geval van een samengestelde zin zowel de hoofd- als bijzin).

NB: agreement kan alleen geanalyseerd worden als de zin zowel een subject als een werkwoord bevat. elliptische zinnen en imperatieveen kan je dus niet coderen op agreement.

is er sprake van subject omissie? codeer dit als VERROR en specificeer de fout in de %err tier.

Codeer congruentie in samengestelde zinnen voor zowel de hoofd- als bijzin, *indien het subject en ww fonologisch worden gerealiseerd* ('de mama eet en de babyvogel ook' is dus: VP+ SVA3S+)

Wordt het subject en/of ww niet fonologisch gerealiseerd? Dit is een grammaticale samentrekking. Codeer dit in de &com tier als: grammaticale samentrekking.

SVA1S+ of – eerste persoon enkelvoud goed/fout gebruikt (ik loop)

SVA2S+ of – tweede persoon enkelvoud goed/fout gebruikt (jij loopt)

SVA3S+ of – derde persoon enkelvoud goed/fout gebruikt (hij/zij loopt)

SVA1P+ of – eerste persoon meervoud goed/fout (wij lopen)

SVA2P+ of – tweede persoon meervoud goed/fout (jullie lopen)

SVA3P+ of – derde persoon meervoud goed/fout (zij lopen)

Complexiteit:

ROOT root infinitive, onvervoegd werkwoord (mama koekje eten, hij de kat aaien). Codeer alleen een ROOT als er een subject in de zin aanwezig is. En let op of er geen sprake is van een elliptisch modaal werkwoord.

DUMMY dummy auxiliary. een hulpwerkwoord zonder lexicaal element + een hoofdwerkwoord met lexicaal element (hij gaat fietsen ipv hij fietst). vaak vormen van /gaan/, /doen/, /zijn/ en /zitten/ (in de context dat er niemand zit).

NB: modale werkwoorden zijn geen dummy auxiliaries. 'hij moet wachten' of 'hij kan fietsen'. dit kan je checken door het hoofdwerkwoord in de zin te vervangen, ipv het hulpwerkwoord: verandert dan de betekenis? zo niet, dan is het hulpwerkwoord een dummy auxiliary.

DITRANS dittransitief werkwoord, dat een direct en indirect object (lijdend voorwerp en meewerkend voorwerp) uitdeelt, bijv. jan geeft de bal aan marie.

Fouten:

VERROR	werkwoordgerelateerde fouten, bijv. omissies van een subject, object of ander argument, volgordefouten of een ongrammaticale zin. Markeer met [*] en specifieer de fout in een %err tier. Agreement-fouten markeer je niet, die worden al meegenomen in een andere code. Bijvoorbeeld: *CHI: dan gaat (d)ie achteraan [*] %maa: VP+ SVA3S+ DUMMY VERROR %err: omissie object
PAST+	verleden tijd goed, in een zin met een vt-ww en vt-adjectief (en <u>toen kwam</u> hond) of een zin waaruit het duidelijk om een gebeurtenis uit het verleden gaat.
PAST-	verleden tijd fout (* <u>vroeger</u> is mijn lievelingskleur wit).

General remarks for transcribing

In case you do not know how to transcribe something, you should first insert a @Comment line with some information and the signs \$\$. These sections can later on be easily tracked.

Concerning the codes in square brackets (e.g., [/] or [*]): these usually apply only to the immediately preceding word. If the code is meant to apply to several words then you should insert the pointed brackets < >. Otherwise pointed brackets do not need to be used. To give you an example, if the transcript says "I am [x 3]" it means the speaker actually said: "I am am am". However, if the transcript says "<I am> [x 3]", then the speaker originally said "I am I am I am".

Concerning written material within square brackets: Comments on the main tier line following [% do not have to correspond to CHAT's minimum requirements; thus, they could also contain Umlaute, numbers, etc. Correct replacement forms after [: however need to correspond to CHAT standards because they replace other material and will be analyzed.

Minimum standards for CHAT files

CHAT only recognizes three utterance terminators: period (.), question mark (?) and exclamation mark (!). These must appear only once in a main tier line, namely at the end. Each main tier line has to end with one of those three utterance terminators. Besides these utterance terminators only , and ; are allowed on the main tier line.

- Words are separated by spaces.
- The beginning of an utterance should otherwise always be typed in lower case.

Wat is een uiting? Een T-unit (minimally terminable unit, naar Hunt (1965))

= *een main clause + all subordinate clause(s) attached to it.*

Uitingen worden gedefinieerd op basis van grammatica, intonatie en pauzes.

Twee hoofdzinnen zijn twee uitingen:

- *CHI: en toen klom de kat in de boom .
- *CHI: en toen pakte die het geitje .

Uitzondering: als de tweede zin materiaal uit de eerste zin nodig heeft:

*CHI: hij greep de vos en jaagde hem weg .

Een hoofdzin kan uit één woord bestaan en hoeft niet per se een werkwoord te bevatten:

*INT: en wat doe je allemaal op school ?

*CHI: rekenen .

*CHI: leren lezen .

*CHI: leren .

*CHI: in de poppenhoek .

Een hoofdzin met één of meer bijzinnen is één uiting:

*INT: en wat kun je allemaal in de zomer doen ?

*CHI: in het zwembad spelen maar dat is niet leuk want daar komen die waterbeestjes .

Je plakt de hoofd- en bijzin ook aan elkaar als de onderzoeker tussendoor iets korts zegt, zoals 'ok@i', of 'ja@i heel goed' .

*CHI: ik hou van buiten spelen want dat vind ik leuk .

*INT: hm@i hm@i .

ipv:

*INT: ik hou van buiten spelen .

*INT: hm@i hm@i .

*CHI want dat vind ik leuk .

Uitzondering: als er een erg lange pauze zit tussen de hoofd- en de bijzin en de intonatie duidelijk naar beneden gaat aan het eind van de hoofdzin. Deze beslissing neem je op je gevoel.

Opsommingen vormen één uiting, tenzij er een lange pauze tussen de opgesomde elementen zit en de intonatie duidelijk naar beneden gaat.

*CHI: eh@fp nou@i eh@fp emma die [/] eh@fp die is heel erg vriendelijk en melle ook .

Woorden als ja, nee en toch horen bij de hoofdzin.

*CHI: ja ik ben toen gehackt .

Het aan elkaar schrijven van woorden:

Deze woorden heb ik aan elkaar geschreven: erin, eraan, daarop, weleens, dichtbij, vanalles

Getallen

Schrijf je volledig uit, elk woord apart: twee honderd twee, twee en vijftig.

Eigennamen

Schrijf je met een hoofdletter: Euro, Iphone, Xbox, Roblox.

Transcriberen van de interviewer

Wat de interviewer zegt, wordt letterlijk genoteerd, maar hoeft niet getranscribeerd te worden.

Toelichting van bepaalde codes:

Stutters, babbling, non-finished words

- & If the spoken material is acoustically recognizable but does not make sense, e.g. due to babbling, stuttering, stumbling, incomplete words etc., then use & for transcribing:

Stutter:

*CHI: and then I &f &f fell .

Stumbling and retracing:

*CHI: and then I &fet [//] fell .

In the second utterance "fet" is obviously a failed attempt to say "fell" – therefore it has to be marked as retracing as well (see below), dit geldt alleen voor incomplete woorden of klanken

Compound words, phrases, collocations

The question about what can be considered one (single) word is relevant for two reasons:

- The length of utterance in words is a possible measure for the complexity of utterances.
- CLAN will automatically create the %mor line for words it can recognize, i.e. it will supply morphological information about all the words of an utterance.

To make it recognizable for CLAN that something is to be counted as one word use the underscore (_).

-
- Use the underscore for several words which are so closely related that they are considered a fixed phrase or collocation (like one long word). These are expressions such as

- you_know, french_fries, Big_Ben, in_god_we_trust, Bar_Mizwa,

Also, use the underscore for acronyms where each letter is pronounced:

- U_S_A, T_shirt

Acronyms which can be pronounced as a single word (e.g., *UNO*, *NATO*) are transcribed as normal words.

The decision to mark something as a phrase with _ may be made on a speaker-by-speaker basis: if you feel that a certain speaker has a favorite phrase and uses it so often that for her/him it constitutes a fixed phrase, you may use _ .

Because such decisions will strongly influence the results for measuring the length of an utterance the underscore should be moderately used. In case of uncertainty, please mark the utterance with @Comment: \$\$\$ (see above).

Spelling, Acronyms

- @I When something is spelled out, append @I to each single letter.

-
- For words which consist of individually pronounced letters:

U_S_A etc. (see previous page)

In case spelling takes place in the second language, an @s to indicate codeswitching should be added (with @s last, yielding @I@s). Titles such as Dr. or Mrs. have to be written out: Doktor, Missus

Incomplete, contracted and dialectal forms

() When speaking rapidly speakers often drop sounds. In the transcript these sounds appear in parentheses to restore the complete word:

(ei)n, (den)n, he (ha)s, I (woul)d

There is no need to use parentheses sparingly, they may appear often!

[:] Non-standard forms and dialectal forms cannot be automatically recognized and annotated by CLAN. Therefore, they are transcribed as heard and repeated in the complete, standard form in square brackets:

*CHI: en toen heeft ie [: hij] (ee)n vis gegeten .

*CHI: whyncha [: why don't you] just be quiet !

Uh, Eh, Ehm

eh@fp filled pause Pauses filled with expressions such as "uh", "oh", etc. are transcribed as eh@fp no matter what was said. E.g., "De moeder ehmmmmm vliegt weg" becomes

*CHI: de moeder eh@fp vliegt weg.

Acoustically difficult or unintelligible material

[?] If you think you recognize a word but are not entirely sure, e.g., because the word was not clearly pronounced then you should mark the word as unclear [?]:

*CHI: I have two [?] brothers .

[=?] If you are not sure whether the one or the other was said then you type into [=? Text] the alternative transcription:

*CHI: we want <one or two> [=? one too].

*CHI: I habe zwei [=? drei] Brüder.

xxx If something is completely unintelligible, e.g., because it was too soft-spoken or due to noise in the background then you transcribe with xxx.

*CHI: and then xxx I fell.

*CHI: xxx.

xxx either implies one/several unintelligible words or an entire utterance.

Nonverbal sound events

If someone cries, coughs, or sneezes we can distinguish whether someone does so during speaking or instead of speaking.

&= Added if sound events occur instead of speaking.

*CHI: *CHI: &=coughs .	&=pants .
---------------------------	--------------

Within a main tier line: repetitions and reformulations

[/] Repetition *Simple Repetition*

The [/] symbol is used in those cases when a speaker begins to say something, stops and then repeats the earlier material without change:

*CHI: <I wanted> [/] I wanted to ask her .

In case there is a pause marked by (.), the utterance may nevertheless be considered a simple repetition, e.g. :

*CHI: he is [/] (.) is pausing .

Don't mark something as a repetition if it is repeated purely for rhetorical effect, such as:

*CHI: I was very very angry !

Also, don't mark something as a repetition if the material following [/] is different from the material preceding [/], for example when there is an eh@fp between the repeated word(s). In these cases, use the retracing symbol instead (see below):

*CHI: I [//] eh@fp I wanted to ask her.

[x N]	Multiple Repetition	<i>Multiple repetitions</i> If the same word is repeated several times, instead of transcribing
		<p>*CHI: <it's it's it's> [/] it's like a dog .</p> <p>use the [x N] marker, with N being the number of times the word is repeated:</p> <p>*CHI: it's [x 4] like a dog .</p> <p>This form indicates the fact that a word has been repeated four times. The [x N] marker should be used for a minimum of 3 repetitions.</p>

Within a main tier line (continued)

[//] Retracing *New attempt for the same utterance*

This symbol is used when a speaker starts to say something, stops, repeats the basic phrase, changes the syntax but maintains the same idea. In retracing with correction, it is necessarily true that the retraced material is different from what follows the retracing symbol:

*CHI: <I wanted> [//] eh@fp I thought I wanted to ask Margie .

Retracing with correction can combine with retracing without correction, as in this example:

*CHI: <the fish is> [//] the [/] the fish are swimming .

[///]	Reformulation	<i>Full and complete reformulation</i> Sometimes retracings involve full and complete reformulations of the message without any specific corrections: *CHI: <all of my friends had> [///] eh@fp we had decided to go home for lunch . Please remember to put <> around the material which is dropped, i.e. reformulated. When none of the material being corrected is included in the retracing, it is better to use [///] than [/].
-------	---------------	--

At the end of a main tier line: trailing off and interruptions

+...	Trailing off	The trailing off or incompleteness marker is the terminator for an incomplete, but not interrupted, utterance where the speaker is not aware of the fact that the utterance is incomplete. Therefore, trailing offs cannot normally be followed by a self-completion (+,). Trailing off thus occurs when speakers
------	--------------	--

- shift attention away from what they are saying, sometimes even forgetting what they were going to say
- do not finish an utterance but feel that it is finished.

Trailing off does not occur when a speaker can't find a word and another speaker completes the utterance (scaffolding). This is marked with +/. and +, (see below).

When a speaker completes an utterance and only the intonation indicates that the words are left "hanging in the air", transcribe it as a normal utterance (with a period, not as trailing off).

Please make sure Word does not change your three separate periods into one sign with three periods.

+/.	Interruption	The interruption symbol is used for an utterance that is incomplete because one speaker is interrupted by another speaker: *CHI: what did you +/.
		*INT: Mommy . *CHI: +, with your spoon .

Note that the last utterance completes the first and is therefore marked as a self-completion (see next section).

+//.	Self-Interruption	A self-interruption occurs if someone interrupts his own current utterance to continue with a next utterance that does not complete the interrupted utterance. The next utterance is therefore not marked as a self-completion: *CHI: smells good enough for +//. *CHI: what is that ?
------	-------------------	--

If a speaker interrupts his current utterance only to continue with it in a different way, this is marked as a retracing or reformulation (see previous section).

At the beginning of a main tier line: continuations

+ ,	Self-Completion	The self-completion symbol is used at the beginning of a main tier line to link an utterance to a preceding main tier line of the same speaker.
-----	-----------------	---

The symbol is used:

- After an interruption, when the interrupted speaker completes the interrupted utterance (see Interruption above)

++	Other-Completion	The ++ symbol marks "latching" or the completion of another speaker's utterance as in the following example:
----	------------------	--

*CHI: if Bill had known +...

*INT: ++ he would have come .

It also marks scaffolding, thus if one speaker misses a word and another speaker helps in constructing the utterance:

*CHI: that must have been in +/.

*INT: ++ Berlin .

The preceding utterance of the other speakers ends with +... if the other speaker is not aware of the fact that his/her utterance is incomplete. Otherwise it ends with +/, such as in the case of scaffolding.

Errors

[*]	Morphological and syntactic errors should be marked with [*]. If the error is not confined to one word, pointed brackets (< >) need to be used to mark the „area“ within which the error occurs:
-----	--

CHI: De hond heeft de poes gepakken [: gepakt] []

CHI: De hond <eten wilde vis> []

Pauses

(.)	Pauses between words are marked with (.), (..) or (...) depending on their length
-----	---

KEUZEN EN CONVENTIES

- (1) Van gereduceerde woorden of uitingen die niet afwijken van de target, worden de niet-gerealiseerde klanken op de Target tier tussen haakjes gezet, zoals 'e(en)s' of 'hou(d)' of '(wa)t is dat', omdat de woorden in de codeerfase door de morphological tagger dan juist worden herkend (de morphological tagger negeert de haakjes).
- (2) Vragen worden gekenmerkt met een ? aan het einde van de uiting; imperatieven worden gekenmerkt met een ! aan het einde van de uiting. Onder imperatieven worden alleen imperatieven met eenzelfde vorm als de stam van het werkwoord verstaan, omdat het lastig te controleren is of infinitieven een imperatieve functie hebben. De consequentie hiervan is dat de negatieve gebiedende wijs meestal niet als zodanig is aangeduid. Bij twijfel over een imperatieve vorm (zoals bij een geïsoleerd 'maak' dat bijvoorbeeld ook een eerste persoonsvorm kan zijn) wordt de uiting niet gekenmerkt met een !.

Appendix B
Procedure for second check of the transcriptions (in Dutch)

- Controleren of achter elke ja/nee een @i staat.
- Ja's en nee's zoveel mogelijk plakken aan de T-unit waar ze bij horen.
- Vishaken < > voor retraces [//] en [///]
- Subject/object omissies nalopen.
 - o Gebruik 'rich interpretation' (volwassen vorm van een kinderzin)
- 1x een VP markering in samengestelde zinnen.
- Alleen SVA markeren als er een subject + werkwoord in de zin staat. Niet annoteren als subject of werkwoord fonologisch niet gerealiseerd is.
 - o Codeer grammaticale samentrekkingen in %com.
- Topicalisatie niet als VERROR rekenen. Bijv: 'een nieuw huis hebben wij in Polen'
 - o Het is niet fout, maar pragmatisch vreemd.
 - o Noem het in %com: topicalisatie
- Uitingen met trailing off nooit annoteren.
- Verleden tijd afleidbaar uit context ook annoteren met PAST.
- 'Gaan' als dummy nalopen: is het zelfstandig gebruikt? Geen dummy. Is het toekomende tijd? Ook geen dummy.
- Vocabiel aanduiden met @v.
- Topic drop niet annoteren als VERROR, maar noemen in %com: topic drop.
- Elke werkwoordsgerelateerde fout = VERROR. Dus 2x VERROR in een zin kan ook.
- ROOT nalopen: annoteer alleen als ROOT als er een subject wordt uitgedrukt.
Verwar het niet met zinnen waarin het hulpwerkwoord elliptisch kan zijn.