Jaargang 27, Special Issue, november 2022

Stem-, Spraak- en Taalpathologie

Special issue ter ere van het emeritaat van professor Steven Gillis

rijksuniversiteit groningen / bibliotheek

Stem-, Spraak- en Taalpathologie

Vol. 27, 2022, Special Issue, pp. i-iii
Special Issue for Steven Gillis

Gepubliceerd online: 5 november 2022

University of Groningen Press
https://doi.org/10.21827/32.8310/2022-SG-i

Introduction

On October 1st, 2022 Steven Gillis retired after a long and rewarding career as a Professor at Antwerp University in the Department of Linguistics and as head of the psycholinguistics section of CLiPS, the Centre for Computational and Psycholinguistics. On the occasion of Professor Gillis' retirement it was decided to produce a special issue of Stem-, Spraak- en Taalpathologie to honour his achievements. This special issue is a collection of research papers which are directly or indirectly related to or inspired by the type of work that Steven has carried out during his career.

Throughout his professional life, Steven Gillis has been a very active researcher and promotor of research. He acquired funding and was Principal Investigator of a total of 51 research projects hosted at Antwerp University. At the moment, he has a total of 239 publications: articles in international peer-reviewed journals, book chapters and conference proceedings. These cover a wide range of research topics in phonetics, phonology, morphology and syntax. But very fundamentally his main focus has always been on first language acquisition broadly conceived, including language development disorders. What stands out, are his studies on the speech and language characteristics of children with hearing impairment and their relationship to auditory support of these children by means of cochlear implants, conventional hearing aids and more recently brainstem implants. In this domain, Steven Gillis' research has always been at the forefront of technology. His research on children with a cochlear implant for instance was one of the first in the world on children who were implanted at a very young age. The same forefront role applies to his studies on children with a brainstem implant, which are still very few and far between. This has provided unique insights in the benefits of technological advances to the speech and language development of children with hearing impairment.

Four contributions in this special issue focus on language development disorders in children. Three of them relate to hearing impairment.

Odijk & Faes investigate the development of the place of articulation feature in children with a brainstem implant (ABI). The results show that the coronal place of articulation was most used in children's speech, except for one ABI child who showed a preference for labials. It was also found that the labial place of articulation was produced more accurately than the coronal place of articulation. It was also observed that incorrect productions of place of articulation often represented omissions rather than substitutions by another place of articulation. This study suggests that ABI children clearly benefit from their device, but still have a long way to go to catch up with their peers.

Genette & Verhoeven study the dynamic aspects of pitch in prominence realisation in children with typical hearing as compared to children with hearing impairment (conventional hearing aid vs. cochlear implant). It is found that prominence realisation in children with a conventional hearing aid is very similar to that of children with typical hearing. In children with a cochlear implant, however, the pitch contours associated with prominence are less dynamic than in the other groups. This is taken as an indication of more monotony in their speech.

Gary Morgan, finally, presents a much-needed reflection on the different approaches that have been taken in the study of deafness and raises debate regarding the medical approach (importance of screening and early implantation) vs. the sociological approach (recognition of variation between children and the right to sign language) in the study of language development of deaf children over the last thirty years.

A fourth contribution focuses on developmental language disorder (DLD). Boerma, Wijnen & Blom studied the performance and predictors in children with DLD after specialist educational provision was discontinued at the age of 7 - 8 years. It was found that the language and reading skills of these children are often poor and suggest that this group remains vulnerable with continued attention being required.

From Steven's publication list it is clear that he has always had an interest in the acquisition of morphological systems. This is, for example, illustrated by his work on the effect of morpheme boundaries on the syllabification of words, the acquisition of diminutives and the acquisition of verb inflection in Hebrew.

Colleagues who share Stevens' interest in morphology, are Tribushinina & Mak, who present a study on the acquisition of pronominal gender in Russian, a morphologically rich language. Their study addresses the question whether bilingual children acquire pronominal gender equally successfully as monolingual Russian children with and without developmental language impairment.

A second morphologically-inspired contribution to this special issue is Dressler, Mattiello, Korecky-Kröll, Noccetti, Dabašinskienė, Kamandulytė-Merfeldienė & Kazakovskaya, who study parallels in the use of diminutives in child-directed speech and pet-directed speech in German, Italian, Lithuanian, Russian and English. Their study illustrates, among other things, how diminutives are used to express emotions and familiarity.

Besides an interest in language development disorders and morphology, Steven has always been fascinated with spelling. For example, he studied the effect of spelling knowledge on intuitive syllabification. In this special issue, this theme is represented by a contribution of De Bree, Lammertink, Van Witteloostuijn & Rispens, who investigated to what extent spelling outcomes of children with dyslexia and DLD are related to literacy and/or oral language difficulties. The results indicate that both groups of children show spelling difficulties, but these are most outspoken in children with DLD who have additional reading difficulties. The paper argues for intensive spelling intervention and for support of both literacy skills and oral language in children at risk of spelling difficulties.

Another common theme in Steven's research that should definitely be mentioned as well, is his concern with reliable research methodology and a judicious use of speech and language corpora. This interest in methodology and corpus use is clearly shared by the authors of the three remaining contributions.

Sijyeniyo, Hartsuiker & Bernolet conducted a norming study amongst Dutch children to investigate whether they correctly interpret a set of 208 drawings, developed for the study of transitive sentences (with some drawings having an active interpretation and others a passive one).

Haim & Ravid compiled a corpus of informative and expository texts of Hebrew-speaking students and adults to study the (evolution in) the complexity of their vocabulary and (morpho) syntax.

Finally, Swerts, Gu & Boerrigter go further than speech, language, reading and writing by studying gesture. This contribution investigated to what extent head gestures of children "correctly" support children's attitudes to objects that they may desire or not.

Acknowledgements

We would like to thank the editorial committee of Stem-, Spraak- en Taalpathologie who agreed to publish this special issue and who facilitated the production of this volume. Also a big thank you to Frits van Brenk who typeset all the submitted articles with great efficiency and with stunning result. We would also like to thank all the reviewers of the submitted articles, who have done a marvelous job in providing feedback on the research presented in this issue. Finally, we would also like to thank all the contributing authors: the wide range of research topics and methodologies have all added up to a great tribute to Steven Gillis as an inspiring academic.

We would like to wish Professor Gillis a long, active and satisfying retirement.

De redacteuren van dit nummer

Jan de Jong[1], *Jolien Faes*[2], *Hanne Kloots*[2], *Jo Verhoeven*[2,3]

[1]Universitetet i Bergen, Institutt for biologisk og medisinsk psykologi, Bergen, Norway
[2]University of Antwerp, Computational Linguistics & Psycholinguistics Research Center (CLiPS), Antwerp, Belgium
[3]City, University of London, Department of Language and Communication Science, London, UK

STEM-, SPRAAK- EN TAALPATHOLOGIE

Gepubliceerd online: 5 november 2022
© 2022 De auteur(s)
University of Groningen Press
https://doi.org/10.21827/32.8310/2022-SG-1

Vol. 27, 2022, Special Issue, pp. 1-23
Special Issue for Steven Gillis

Discontinuation of specialist educational provision for children with DLD in the Netherlands: performance and predictors

Tessel Boerma[1], Frank Wijnen[1], & Elma Blom[2]

[1] *Department of Languages, Literature and Communication & Institute of Linguistics OTS, Utrecht University*

[2] *Department of Education & Pedagogy, Development and Education of Youth in Diverse Societies, Utrecht University*

Samenvatting

Achtergrond. Taalontwikkelingsstoornis (TOS) wordt gekenmerkt door grote heterogeniteit, maar er is nog weinig bekend over de variatie in specialistische onderwijsondersteuning voor kinderen met TOS. Dit longitudinale onderzoek bestudeerde de taal- en leesvaardigheid van kinderen met TOS bij wie onderwijsondersteuning op 7- of 8-jarige leeftijd beëindigd werd en kinderen met TOS bij wie dit werd voortgezet, ten opzichte van typisch ontwikkelende kinderen. Daarnaast onderzochten we talige, cognitieve en omgevingsfactoren als voorspellers voor de beëindiging van specialistische onderwijsondersteuning.

Methode. Eentalige en meertalige kinderen met (n=120) en zonder (n=113) TOS namen deel aan het onderzoek. Gestandaardiseerde instrumenten werden gebruikt om de proportie van kinderen per groep met zwakke taal- en leesvaardigheid op 7- of 8-jarige leeftijd te beschrijven. Binaire logistische regressie werd uitgevoerd om de beëindiging van specialistische onderwijsondersteuning te voorspellen. Voorspellers waren talige en cognitieve vaardigheden op 5- of 6-jarige leeftijd, groei in taalvaardigheid, onderwijsniveau van ouders en meertaligheid.

Resultaten. Totaal had 25% van de kinderen met TOS geen specialistische onderwijsondersteuning meer op 7- of 8-jarige leeftijd. Meer meertaligen dan eentaligen behielden onderwijsondersteuning, hoewel dit verschil net niet significant was. Kinderen met TOS, zowel met als zonder onderwijsondersteuning, scoorden vaker zwak op de gestandaardiseerde instrumenten dan typisch ontwikkelende kinderen. Kinderen met TOS met specialistische onderwijsondersteuning presteerden vaker zwak dan kinderen met TOS zonder ondersteuning. Onderwijsniveau van ouders en morfologische vaardigheden, zowel op baseline als groei, voorspelden de beëindiging van onderwijsondersteuning.

Correspondentieadres:
Tessel Boerma
Utrecht University
Dept. of Languages, Literature and Communication
Trans 10, 3512 JK, Utrecht, The Netherlands
E-mail: t.d.boerma@uu.nl

Conclusies en implicaties. Specialistische onderwijsondersteuning bleek voor veel kinderen met TOS binnen een relatief kort tijdsbestek beëindigd te zijn. Hoewel deze kinderen sterker presteerden dan leeftijdsgenoten met TOS die ondersteuning behielden, was hun taal- en leesvaardigheid vaak zwak. Deze groep blijft dus kwetsbaar en behoeft aandacht. Vervolgonderzoek naar besluitvorming omtrent specialistische onderwijsondersteuning is nodig om de rol van onderwijsniveau van ouders en morfologische vaardigheden in de beëindiging van ondersteuning te begrijpen.

Abstract

Background. Developmental Language Disorder (DLD) is characterized by large heterogeneity, but little is known about the variation in specialist educational support that children with DLD receive. This longitudinal study compared the language and reading performance of children with DLD for whom provision had been discontinued at age 7 or 8 years and children with DLD for whom it was continued, relative to typically developing (TD) controls. Additionally, we investigated linguistic, cognitive and environmental predictors of the discontinuation of provision.

Methods. Monolinguals and bilinguals with (n=120) and without (n=113) DLD participated. Standardized measures were used to examine the proportion of children per group with poor language and reading ability at age 7 or 8. Binary logistic regression was conducted to predict the discontinuation of specialist educational provision, including language and cognitive skills at age 5 or 6, language growth, parental education and bilingualism.

Results. Specialist educational provision was discontinued for 25% of the children with DLD. More bilinguals than monolinguals received continuing support, although this difference fell just short of significance. Children with DLD, both with and without specialist provision, performed more poorly on the standardized instruments than TD children, and the children with DLD with provision more often scored poorly than those without provision. Parental education, as well as children's baseline scores on and growth in inflectional morphology, predicted the discontinuation of specialist educational support.

Conclusions and implications. Within a relatively short time period, specialist educational provision was discontinued for many children with DLD. Although these children perform better than peers with DLD with continuing specialist support, their language and reading skills are often poor. This group thus remains vulnerable and requires attention. Future research into decisions about specialist educational provision is needed to understand the role and relevance of inflectional morphology and parental education in the (dis)continuation of provision.

Introduction

Children with a Developmental Language Disorder (DLD) experience severe language difficulties which affect development and behavior in language and other domains of functioning, as well as their future life prospects (Curtis, Frey, Watson, Hampton, & Roberts, 2018;

Dubois et al., 2020; Eadie et al., 2018). There is a lack of clarity about the etiology of these language difficulties (Bishop, 2009), but there is consistent evidence that DLD is persistent (Botting, 2020; Clegg et al, 2005). The group of children with DLD is characterized by large heterogeneity (Bishop, Snowling, Thompson, Greenhalgh, & Catalise-2 consortium2017), for example in terms of the language domains that are affected, and the severity of the problems. Another factor that varies between individuals with DLD is the amount of support they receive (Clegg et al., 2005; Dockrell & Lindsay, 2008; Dockrell, Ricketts, Palikara, Charman, & Lindsay, 2019). Some individuals with DLD may receive intensive support throughout an extensive part of their lives, for example through specialist education services, whereas others receive such support only during a short period of time (Clegg et al., 2005; Dockrell & Lindsay, 2008; Dockrell et al., 2019). While the phenotypical heterogeneity of DLD is widely acknowledged, little is known about the factors that influence these differences in (educational) provision. For example, is short-term provision associated with children with ‘transient’ or ‘resolved’ DLD, with language outcomes indistinguishable from typically developing (TD) peers (as in Bishop & Edmundson, 1987)? Or do these children benefit from compensatory abilities or circumstances, such as high intellectual functioning or socioeconomic status?

Investigating these issues is important, as discontinuation of specialist educational support could present a risk if these children continue to struggle with language and additionally, develop reading difficulties and/or socio-emotional problems. Prognosis regarding the provision of specialist support is relevant for children and parents, but also for the professionals working with the children, and could provide insight into possible compensatory mechanisms that enable children with DLD to develop and learn without additional support. Therefore, the threefold aim of this study is to 1) establish for how many children with DLD in our Dutch cohort specialist educational support is discontinued at age 7 or 8 years old, 2) investigate how these children perform on standardized and norm-referenced language and literacy measures at this age, and 3) explore which factors predict the (dis)continuation of specialist educational provision at this age, including both language and cognitive skills two years earlier, as well as environmental factors.

Difficulties of children with DLD

The language difficulties of children with DLD can surface in all language domains. Although the symptoms of DLD can be manifold, difficulties with inflectional morphology are often considered highly characteristic of DLD (Leonard, 2014), and morphosyntactic errors have been the basis for clinical markers that are considered important for diagnosing DLD (Leonard, 2014; Rice & Wexler, 1996). Next to inflectional morphology, children with DLD are also known to often have weak syntactic and phonological skills, as for example reflected by poor performance on, respectively, sentence (Riches, 2012) and nonword repetition (Graf-Estes et al., 2007; Schwob et al., 2021) which have both also been proposed as clinical markers (Conti-Ramsden, Botting, & Faragher, 2001). Vocabulary is sometimes seen as a relative strength of children with DLD, but weaknesses relative to TD peers have been found in this domain as well (Rice & Hoffman, 2015). Narrative skills are also often reported to be impaired in children with DLD (Duinmeijer, de Jong, & Scheper, 2012; Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016).

Although the language difficulties are of primary concern, children with DLD tend to perform less well than their TD peers in other domains. Co-morbidities with other neurodevelopmental disorders such as dyslexia (McArthur, Hogben, Edwards, Heath, & Mengler, 2000), autism spectrum disorder (Bishop, 2010), attention disorders (Mueller & Tomblin, 2012) and motor difficulties (Webster et al., 2006) are frequently observed, which is in line with comorbidity rates in other developmental disabilities (Gilger & Kaplan, 2001). More generally, meta-analyses have shown that children with DLD tend to perform lower than their TD peers with respect to attention and executive functioning (Aljahlan & Spaulding, 2021; Ebert & Kohnert, 2011; Pauls & Archibald, 2016; Vugs, Cuperus, Hendriks, & Verhoeven, 2013), as well as intellectual functioning (Gallinat & Spaulding, 2014). This does, however, not hold for all children (e.g., Blom & Boerma, 2020). Problems of children with DLD, both within and beyond the domain of language, are highly variable.

Persistence of DLD

DLD is a disorder which has lifelong consequences. Long-term longitudinal studies show that the majority of children with DLD experience difficulties into adolescence and adulthood, both in and beyond the domain of language (Beitchman et al., 1996; Bishop & Edmundson, 1987; Botting, 2020; Clegg et al., 2005; Conti-Ramsden, Botting, Simkin, & Know, 2001; Dubois et al., 2020; Johnson et al., 1999; Scarborough & Dobrich, 1990; Snowling et al., 2006; Snowling et al., 2015; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998; Tomblin, Zhang, Buckwalter, & O'Brien, 2003; Zambrana et al., 2014). For example, Botting (2020) analyzed follow-up data from individuals with DLD at age 24 years and found that only 26% obtained language scores in the average range. In the study by Johnson and colleagues (1999), including individuals at age 19 years, this was only 13%. At a younger age, investigating children between 3 and 8 years old, Snowling and colleagues (2015) report that 16% had resolved DLD, as defined by language scores in the average range. These relatively low percentages indicate that language problems are difficult to overcome and reflect the stability of language development trajectories across childhood, especially after age 4 years, as is also observed in TD children (Bornstein, Hahn, Putnick, & Suwalsky, 2014) and child population samples (McKean et al., 2017a; Norbury et al., 2017).

Given the persistent consequences of DLD, the common goal of professionals supporting children with DLD is to maximize a "child's functioning, activity, well-being and participation, in both education and socially" (Ebbels, McCartney, Slonims, Dockrell, & Norbury, 2019; p.5). When deciding on which type of support is best suited for a child, individual characteristics which have been found to predict whether the problems of a child with DLD persist over time, are thus highly relevant to consider. Early language ability is one of the strongest predictors for language outcomes later in life (McKean et al., 2017b). Research shows that the lower a child scores on standardized language measures, and thus the more severe the impairment may be, the more likely this child is to have persistent language problems and, moreover, develop reading difficulties when growing up (Snowling, Duff, Nash, &

Hulme, 2015). Particularly children with low scores on both expressive and receptive language tests are at risk for poor outcomes later in life, both in the domain of language (Bishop & Edmundson, 1987; Zambrana, Pons, Eadie, & Ystrom, 2014), as well as socio-emotional and behavioral development (Beitchman et al., 1996; Snowling, Bishop, Stothard, Chipchase & Kaplan, 2006). Other research also points to narrative skills as an important predictor (Bishop & Edmundson, 1987). The severity of DLD and types of problems are thus informative about prognosis and co-occurring problems beyond language, and are, as indicated by a survey study in the United Kingdom (Dockrell, Lindsay, Letchford, & Mackie, 2006), therefore often taken into account when decisions on specialist educational provision are made.

However, next to characteristics which are linguistic in nature, there are also cognitive and environmental factors which have been found to predict the persistence of DLD. In a meta-analysis on language outcomes in late talkers, Fisher (2017) showed that parental socioeconomic status, next to early expressive vocabulary and receptive language, was significantly related to a child's later language skills. Moreover, a study by Botting (2005) also points to the importance of nonverbal intellectual functioning. The fact that a discrepancy between non-verbal and verbal abilities was reported to be one of the main criteria for admission to specialist provision in the United Kingdom (Dockrell et al., 2006), highlights the importance of individual characteristics beyond the domain of language.

Specialist educational provision of children with DLD in the Netherlands

Allocation of support to children with DLD in schools or private clinical settings is important, because it helps reducing barriers in communication and learning caused by severe and persistent language problems. However, not all children may receive the support they need. Comparing support provided to children with DLD and Autism Spectrum Disorder (ASD), Dockrell et al. (2019) found that children with ASD are more likely to receive support than children with DLD, independent of language, literacy, cognitive scores, and behavior. The study by Dockrell and colleagues is situated in the United Kingdom, and while specialist educational provision for children with disabilities varies from country to country (Wood & Bates, 2020) and provision of services to children with DLD is dependent on national policies and governance (Law, McKean, Murphy, & Thordardottir, 2019), the issue of what drives specialist educational support is relevant across countries. For this study, we specifically looked into *discontinuation* of specialist educational provision of children with DLD. This study is situated in the Netherlands and, before turning to the present study, we briefly describe the Dutch support system insofar as it is relevant to DLD.

Special needs education in the Netherlands is organized in 'cluster-schools' for visually impaired or blind children (cluster 1), children who are deaf, hard of hearing, or have speech and/or language difficulties (cluster 2), children who have cognitive or physical disabilities (cluster 3), and children with behavioral or psychiatric problems (cluster 4). Cluster 2 school teams consist of special educational needs-teachers, remedial teachers, speech and language therapists, and child psychologists. In 2021, a total of 71,605 children and adolescents attended special primary and secondary education; 6,613 were enrolled in cluster 2 primary education and 1,589 in cluster 2 secondary education (Nederlands Jeugdinstituut, 2022), many of whom are diagnosed with DLD. However, more than 50% of the children with DLD take part in mainstream education and receive additional specialist support there (Koninklijke Auris Groep 2017, as cited in Gerrits et al., 2019). Many of these children enrolled in mainstream education also visit speech language therapy practices after school hours (Gerrits et al., 2019).

The law states that parents have the right to choose a school for their child, but specialist educational provision is also dependent on eligibility. Whether or not a child is eligible, either for special education or mainstream education with additional specialist support, is determined by a committee of inquiry (Commissie van Onderzoek 'CvO') of the organization that provides specialized education and support, in agreement with parents and school (Stichting Siméa, 2014). The decision is based on a development perspective plan, drafted by the school, that motivates the application for specialized support. It considers a child's learning capacity and skills, socio-emotional development, communicative self-reliance in educational situations, academic perspective, goals for the child, as well as individual risk and protective factors (Stichting Siméa, 2017). In addition, a child's difficulties need to be sufficiently severe as indicated by: 1) a diagnosis of DLD provided by a multidisciplinary assessment team or suspected DLD if children are still very young, 2) no effects of routine speech and language therapy, and 3) severe problems in one or more of the following four language areas: speech, grammar, semantics, pragmatics, as demonstrated by an overall score of below -2 standard deviations (SD) on a norm-referenced language assessment battery, or -1.5 SD on at least two of the four areas, or -1.3SD on at least three out of the four areas (Stichting Siméa, 2017). The committee of inquiry also determines the duration of specialist educational provision.

Present study

Some children with DLD receive specialist educational provision throughout primary education (and beyond), while for others additional specialized support is discontinued. Little is known about these two groups of children. The main aim of this study was to fill this gap.

Our first aim is to establish how many children with DLD in the cohort that we followed did not receive specialist educational provision anymore at age 7-8 years. In addressing this issue, it is relevant to distinguish between monolinguals and bilinguals. The reason is that, in the Netherlands, bilingual children are overrepresented in special education for children with DLD (Smeets, Driessen, Elferink, & Hovius, 2009), which may suggest that monolingual children are less likely to receive specialist educational provision than bilinguals. Note that specialist educational provision may not only have positive consequences: it could also mean that bilingual children are more likely to be considered at risk than monolinguals and, as a result, are deprived of mainstream education.

The second research aim is to investigate how 7- and 8-year-old children with and without specialist educational provision perform on standardized and norm-referenced language and literacy measures, and to determine if patterns are the same for monolingual and bilingual children. It may be expected that the children who do not receive specialist educational provision are more likely to score within the normal range on language measures than their peers with educational provision, and that their language problems are to some extent more transient (Bishop & Edmundson, 1987). However, whether this is indeed the case, is currently unknown. It is important to examine this, as the children without specialist educational provision may run the risk of falling through the cracks if their language skills are still weak. This is not only relevant for language outcomes but also for literacy skills, because children with weak language skills at earlier ages are likely to develop literacy problems at age 7 (McKean et al., 2017b), and children with DLD show consistently lower reading outcomes than TD peers (Catts, Bridges, Little, & Tomblin, 2008).

The third research aim of the current study is to retrospectively explore factors that may predict (dis)continuation of specialist educational provision at age 7-8 years, focusing on linguistic, cognitive, and environmental variables. Children with DLD with higher language outcomes at age 5-6 years or who show a relatively steep growth in language may be more likely to do without specialist educational provision at age 7-8 years, as they may have a less severe and persistent disorder (McKean et al., 2017b). As explained above, factors related to children's functioning in a school environment and their future academic prospects are also considered in the decision to (dis)continue educational provision, in addition to severity of the condition. Therefore, those areas in language that are thought to be crucial for academic performance, such as vocabulary and narrative skills (Bishop & Edmundson, 1987; Dickinson & Tabors, 2002), could have a higher predictive value than the areas that are affected most by DLD, such as morphosyntax. Relatedly, competencies outside the domain of language, such as sustained attention and learning capacity (Dockrell et al., 2019; West, Shanks, & Hulme, 2021), or environmental variables related to the amount of support that can be provided at home by parents (Fisher, 2017), could be significant predictors.

Methods

Participants

The data for the present study were collected within the longitudinal research program called 'Cognitive Development in the Context of Emerging Bilingualism' at Utrecht University. Participating children were followed over the course of two years and were tested three times with yearly intervals. At wave 3, the children were around 7 or 8 years old. The sample of the present study includes all children for whom data from the first and third wave were available. From this sample, four children with DLD were excluded, because they transferred to special education for children with intellectual disability (cluster 3) during the data collection. One child with DLD was furthermore excluded due to hearing impairment, and another child with DLD was excluded because we had no information on the specialist educational provision at wave 3. Finally, two TD children were excluded, as one was diagnosed with ASD and one with DLD during the course of the study. The final sample included 233 children (see Table 1 for background characteristics).

The children with DLD were recruited via organizations that provide care and education services for children with communication problems (mainly Royal Dutch Kentalis and

Table 1: Background characteristics of the groups of participants

			Sex	**Age in months at wave 3**	**Nonverbal intelligence**	**Parental education**[a]
		N	Girls/Boys	Mean (*SD*)	Mean (*SD*)	Median (range)
DLD	**Monolingual**	88	22/66	94.8 (6.8)	94.2 (17.6)	5.5 (2-9)
	Bilingual	32	10/22	94.7 (8.8)	94.7 (15.3)	5.8 (2-9)
TD	**Monolingual**	43	20/23	94.3 (8.1)	106.5 (15.4)	7.0 (2-9)
	Bilingual	70	39/31	91.5 (7.1)	96.8 (13.8)	5.0 (1-9)

[a] This information was missing for four monolingual children with DLD and seven bilingual typically developing children.

Royal Auris Group). At the start of the longitudinal study, they had all been found eligible for specialist educational support for their language difficulties. This was determined through a standardized protocol (Stichting Siméa, 2017), as described in the introduction, before and independent of their participation in the research. For the bilingual children, a bilingual anamnesis was used and, if possible, testing was done in both languages (following the guideline of Stichting Siméa, 2016). During data collection, relevant changes in, among others, children's eligibility for specialist educational support were monitored through close contact with the schools, parents and other caregivers. The TD children were recruited via regular elementary schools. There were no concerns about their development.

All participating children were born in the Netherlands. A parental questionnaire (*Questionnaire for Parents of Bilingual Children* (PaBiQ); Tuller, 2015) was used to determine whether a child was monolingual or bilingual. Bilingual children had one or both parents who were native speakers of a language other than Dutch and spoke their native tongue with the child throughout an extensive period of the child's life. We measured the percentage of exposure to Dutch before the age of 4 years (bilingual TD: $M = 44.6$, $SD = 10.1$; bilingual DLD: $M = 40.9$, $SD = 11.1$) and percentage of exposure to Dutch at home at wave 1 (bilingual TD: $M = 54.2$, $SD = 13.4$; bilingual DLD: $M = 45.2$, $SD = 16.5$). All bilingual TD children were from Turkish or Moroccan descent, and thus came from the largest immigrant groups in the Netherlands. The bilingual DLD group was more heterogeneous in terms of other languages, but over 70% of the children was also from Turkish or Moroccan descent. Monolingual children had parents who spoke Dutch with them.

Instruments

Parents and/or schools of children with DLD were asked about the specialist educational provision of a participating child at each wave of the study. For the current study, the information from wave 3 will be used to categorize children with DLD in a group with and a group without specialist educational provision. Other instruments are described below, starting with the standardized language and reading measures which are relevant for the second research aim, followed by the instruments which we additionally use in our prediction analysis for the third research aim.

Inflectional Morphology. The subtest Word Formation of the *Taaltoets Alle Kinderen* (TAK [Language Test for All Children]; Verhoeven & Vermeer, 2001) was used to measure Dutch inflectional morphology. In this test, children saw a picture and heard an incomplete sentence. They were asked to finish the sentence, thereby eliciting the plural form of a noun (12 items) and the past participle of a verb (12 items). The maximum score was 24 points. Based on their raw scores and school grade, children can be categorized into norm groups. There are five norm groups for monolingual children ('A' through 'E') and three for bilingual children ('high' through 'low'). For the current study, monolingual children in norm group 'D' and 'E' and bilingual children in norm group 'low' were classified as poor performers.

Function words and sentence patterns. The subtest Sentence Formation of the TAK (Verhoeven & Vermeer, 2001) was used to measure knowledge of Dutch function words and sentence patterns. Children heard a sentence and were asked to repeat this sentence as precisely as possible. Each sentence contained a sentence pattern and a function word which could both be awarded one point if repeated correctly. With 20 sentences in total, children could thus maximally score 40 points. For this test, the same norm groups were used as for the TAK Word Formation subtest.

Receptive vocabulary. The *Peabody Picture Vocabulary Test* (PPVT-III-NL; Schlichting, 2005) was used to test Dutch receptive vocabulary. Children saw four pictures and heard a target word. They were asked to point to the picture which corresponded to the target word. The PPVT contains 204 items which are divided over 17 sets with increasing complexity. In agreement with the official guidelines, a starting set was determined based on a child's age and the test was terminated if the child made nine or more errors in one set. Based on a child's age, raw scores can be converted to norm scores with a mean of 100 and standard deviation of 15. For the current study, children scoring below -1 SD were classified as poor performers. Bilingual norms are not available for this measure.

Word and pseudoword reading. Word reading ability was measured with the *Eén-Minuut-Test* (EMT [One-Minute-Test]; Brus & Voeten, 1999) and pseudoword reading with the Klepel (Van den Bos, Lutje Spelberg, Scheepstra, & De Vries, 1994). Children were presented with a list of 116 words (EMT) or pseudowords (Klepel) which increased in difficulty. They were asked to read as many words as possible in one minute, and as many pseudowords as possible in two minutes. Based on a child's age, the number of correctly read (pseudo)words can be converted to norm scores with a mean of 10 and a standard deviation of 3. A standard score of 6 or lower was considered poor.

Phonological short-term memory. Phonological short-term memory was measured with the Dutch version of the *Cross-Linguistic Nonword Repetition Task* (CL-NWRT; Boerma et al., 2015; Chiat, 2015). Children were asked to repeat a total of 16 nonwords. The nonwords had two to five syllables and were composed of sounds and sound structures which are frequent in many languages in the world. Bilingual children are therefore not disadvantaged due to having had less exposure to a specific language (Boerma et al., 2015). Responses were scored on a whole-item basis, which meant that the maximum score was 16.

Narrative production. Narrative skills were measured with the *Multilingual Assessment Instrument for Narratives* (MAIN; Gagarina et al., 2012; Blom, Boerma, & de Jong, 2020). Children first heard a model story based on a picture sequence. Subsequently, children saw a different picture sequence and were asked to tell their own story. The number of correct plot elements (including the setting, goal, attempt, result, and internal states) in their story were scored and constituted the measure for narrative production. The maximum score was 17.

Sustained attention. Sustained attention was measured with an integrated auditory and visual *Continuous Performance Task* (CPT; see Boerma et al., 2017). Children saw or heard the number '1' or the number '2'. They were asked to press the space bar in response to the number '1', which was the target, and refrain from responding to the number '2', which was the distractor. There were 168 test trials, evenly distributed between targets and distractors. The outcome variable 'd-prime' was used, which is a dual score based on both the ratio hits in response to targets as well as false alarms in response to distractors.

Nonverbal intellectual functioning. The short version of the *Wechsler Nonverbal-NL* (Wechsler & Naglieri, 2008) was used to measure nonverbal intellectual functioning. Two subtests, Matrices and Recognition, were administered with the use of minimal verbal instructions. Raw scores on both subtests can be converted to T-scores and, together, constitute a quotient score with a mean of 100 and a standard deviation of 15.

Parental Education. The *PaBiQ* (Tuller, 2015) was used to gain information on the education level of a child's parents. Education level was measured on a nine-point scale, ranging from 0 (no education) to 9 (university degree). This was filled in for a child's mother and father, and the average of both was taken as outcome variable.

Procedures

The longitudinal research program was approved by The Standing Ethical Assessment Committee of the Faculty of Social and Behavioral Sciences at Utrecht University (file number 22-0098). Informed consent forms were signed by parents of participants. Children were individually tested in a quiet room at school. A native speaker of Dutch administered tasks tapping into language, attention and memory. At each wave of data collection, there were two test sessions of approximately one hour.

Data analysis

Data analysis was done in SPSS, version 26 (IBM Corp., 2016). For the first research aim, we counted the number of children who did (DLD+) and did not (DLD-) have specialist educational provision at wave 3 and assessed whether the groups differed on background characteristics. Using a Pearson's chi-squared (χ^2) test, we furthermore investigated whether the proportion of children in the DLD+ and DLD- groups differed in the monolingual and bilingual group of children with DLD. This is a clean comparison, as these groups did not differ on any of the background variables which were presented in Table 1 (age at wave 3: F(1,118) = .01, p = .94, η^2_p<.001; nonverbal intelligence: F(1,118) = .03, p = .87, η^2_p<.001; parental education: U = 1244.5, z = -.62, p = .54; sex: χ^2(1, N=120) = .47, p = .49).

For the second research aim, we investigated the number of children who performed poorly on the standardized language and reading measures described above. We compared the DLD+, DLD- and TD groups, as well as monolinguals and bilinguals separately. As standardized scores were used, which correct for age, potential age differences between the groups were accounted for. Other background variables are, however, not taken in account and comparisons between the groups are therefore reported here. The monolingual TD children and monolingual children with DLD differed on nonverbal intelligence (F(1,129) = 15.5, p < .001, η^2_p=.11) and parental education (U = 1146.0, z = -3.38, p = .001), with monolingual TD children scoring higher than monolingual children with DLD. The group of monolingual children with DLD included a relatively high number of boys in comparison with the monolingual TD group (χ^2)(1, N=131) = 6.2, p = .01). In the bilingual groups, there were no differences on nonverbal intelligence (F(1,100) = .48, p = .49, η^2_p=.01) and parental education (U = 832.0, z = -1.39, p = .16). The group of bilingual children with DLD included a relatively high number of boys in comparison with the group of bilingual TD children (χ^2)(1, N=102) = 5.3, p = .02). Moreover, there were no differences between the bilingual groups in terms of exposure to Dutch before the age of 4 years (F(1,93) = 2.60, p = .11, η^2_p=.03), but the bilingual TD children received significantly more Dutch exposure at home at wave 1 than the bilingual children with DLD (F(1,92) = 8.07, p = .006, η^2_p=.08).

Finally, in line with the third research aim, we did a binary logistic regression analysis to explore predictors of the discontinuation of specialist educational provision in our sample. Our selection of variables is based on linguistic, cognitive and environmental factors that may be considered when deciding on specialist educational provision for children with DLD in the Netherlands (Stichting Siméa, 2017) and that have been found to predict later outcomes in previous research, as has been described in the introduction. The following variables, all assessed at wave 1, were added as predictors in our model: Dutch inflectional morphology, Dutch knowledge of function words and sentence patterns, Dutch receptive vocabulary, phonological short-term memory, narrative production, sustained attention, nonverbal intellectual functioning, parental education, and bilingualism. Furthermore, we included two variables which were indicative of language growth. We subtracted the raw scores of the TAK Word Formation and TAK Sentence Formation at wave 1 from the raw scores at wave 3. Both difference scores were included in our prediction model. The backward stepwise method was used to report results from the full model with all predictors and the final model which minimized the number of predictors.

Results

Research aim 1. DLD+ vs. DLD-: numbers and background characteristics

Table 2 shows the characteristics of children with DLD who did (DLD+) and did not (DLD-) have specialist educational provision at wave 3. In total, a quarter of the children with DLD did not receive specialist educational support anymore at wave 3. In the monolingual group, this was one-third of the children, while it was only one-eighth of the children in the bilingual group. This difference between the monolingual and bilingual groups fell just short of significance (χ^2)(1, N=120) = 3.6, p = .057).

The DLD+ and DLD- groups did not differ in age at wave 3 (F(1,118) = .84, p = .36, η_p^2=.01), nonverbal intelligence (F(1,118) = .09, p = .77, η_p^2=.001) and sex (χ^2)(1, N=120) = .23, p = .63)), but there were significant differences in parental education (U = 934.0, z = -2.10, p = .04). Parents of children in the DLD- group had, on average, a higher educational level than parents of children in the DLD+ group. The group comparisons showed similar results when analyzing the monolingual and bilingual group separately, although the difference in parental education was not significant in the bilingual group. Percentage of exposure to Dutch before the age of 4 years (bilingual DLD+: M = 39.6, SD = 10.3; bilingual DLD-: M = 50.0, SD = 13.6) and exposure to Dutch at home at wave 1 (bilingual DLD+: M = 42.9, SD = 13.9; bilingual DLD-: M = 60.8, SD = 26.7) were also not significantly different between the groups. However, as there were only four bilingual children in the DLD- group, these results must be interpreted with caution.

Table 2: Children with DLD who did (DLD+) and did not (DLD-) meet the criteria for specialist educational support at wave 3

			Gender	**Wave 3**	**Nonverbal intelligence**	**Parental education**[a]
		N (%)	Girls/Boys	Mean (*SD*)	Mean (*SD*)	Median (range)
DLD+	**Total**	90 (75.0%)	25/65	94.4 (7.7)	94.0 (16.4)	5.5 (2-9)
DLD-		30 (25.0%)	7/23	95.9 (6.2)	95.1 (18.8)	6.3 (3-9)
DLD+	**Monolingual**	62 (70.5%)	46/16	94.5 (7.2)	93.5 (17.1)	5.5 (2-8.5)
DLD-		26 (29.5%)	20/6	95.7 (5.9)	95.7 (19.0)	6.3 (3-9)
DLD+	**Bilingual**	28 (87.5%)	9/19	94.4 (8.9)	95.3 (15.0)	5.8 (2-9)
DLD-		4 (12.5%)	1/3	97.0 (8.6)	91.0 (19.1)	6.0 (3.5-8)

[a]This information was missing for four monolingual children.

Research aim 2. DLD+ vs. DLD- vs. TD: standardized language and literacy measures

Table 3 presents the number of children per group who perform poorly on standardized language and reading instruments at wave 3. For all instruments and for both the monolingual and bilingual group, it can be observed that the proportion of TD children performing poorly on the standardized instruments is lower than the proportion of children with DLD (both DLD+ and DLD-). In addition, more children in the DLD+ group than in the DLD- group perform poorly on the TAK word formation and TAK sentence formation task, with the largest difference on the word formation task. This can be observed in both the monolingual and bilingual group. No clear distinction between the DLD- and DLD+ group is seen on the PPVT, neither in the monolingual nor in the bilingual group. On the reading measures, there are more poor readers in the monolingual DLD+ group than in the monolingual DLD- group, whereas this is not the case in the bilingual group. Again, results in the bilingual group warrant caution given the small sample in the bilingual DLD- group.

Table 3: Number of children per group performing poorly on standardized language and reading measures at wave 3

		TAK word formation[a]	TAK sentence formation[a]	PPVT[b]	EMT[c]	Klepel[c]
		N (%)	*N (%)*	*N (%)*	*N (%)*	*N (%)*
DLD+		64 (71.1%)	72 (80.0%)	21 (23.3%)	37 (41.1%)	36 (40.0%)
DLD-	***Total***	9 (30.0%)	20 (66,7%)	5 (16.7%)	9 (30.0%)	11 (36.7%)
TD		16 (14.2%)	21 (18.6%)	14 (12.4%)	8 (7.1%)	7 (6.2%)
DLD+		47 (75.8%)	48 (77.4%)	6 (9.7%)	30 (48.4%)	29 (46.8%)
DLD-	**Monolingual**	8 (30.8%)	17 (65.4%)	3 (11.5%)	8 (30.8%)	9 (34.6%)
TD		6 (14.0%)	8 (18.6%)	1 (2.3%)	4 (9.3%)	4 (9.3%)
DLD+		17 (60.7%)	24 (85.7%)	15 (53.6%)	7 (25.0%)	7 (25.0%)
DLD-	**Bilingual**	1 (25.0%)	3 (75.0%)	2 (50.0%)	1 (25.0%)	2 (50.0%)
TD		10 (14.3%)	13 (18.6%)	13 (18.6%)	4 (5.7%)	3 (4.3%)

[a] For monolingual children, the number of children scoring in category D and E is counted (lowest 25%). For bilingual children, the number of children scoring in category 'low' is counted (lowest 16%).
[b] The number of children scoring 1 SD below the mean is counted (<85).
[c] The number of children scoring below 7 is counted.

Research aim 3. Predictors of discontinuation of specialist educational provision

Table 4 presents the outcome of the binary logistic regression with which we aim to identify variables that predict the discontinuation of specialist educational support in our sample.

The binary logistic regression showed that our model including eleven predictors was statistically significant from the intercept only model (χ^2)(12, *N*=109) = 37.34, $p < .001$, Nagelkerke R^2 =.43). In this model, scores on the TAK Word Formation task at wave 1 and parental education significantly predicted whether children with DLD would have specialist educational provision at wave 3. Discontinuation of provision was more likely for children with higher TAK Word Formation scores and for children with parents with a higher parental education level. There was also a marginally significant effect of growth on the TAK Word Formation task (i.e., the difference in scores on wave 3 and wave 1). These three predictors were also included in the final model (χ^2)(3, N=109) = 31.88, $p < .001$, Nagelkerke R^2 =.38), using the backward stepwise method. When we ran the analysis with data from only the monolingual children, the final model also included the same predictors.

Table 4: Outcome of binary logistic regression

Model	**Predictors included**	**B**	**S.E.**	Wald	***p***
	Inflectional morphology	-.253	.124	4.124	.042
	Function words and sentence patterns	-.073	.059	1.564	.211
	Receptive vocabulary	.025	.034	.530	.467
	Phonological short-term memory	-.010	.019	.279	.597
Full	Narrative production	-.036	.032	1.209	.271
	Sustained attention	.304	.430	.499	.480
	Nonverbal intellectual functioning	-.002	.019	.011	.917
	Parental education	-.394	.173	5.165	.023
	Growth on inflectional morphology	-.194	.106	3.353	.067
	Growth on function words and sentence patterns	-.045	.054	.687	.407
	Bilingualism	1.320	.905	2.127	.145
	Inflectional morphology	-.362	.088	17.016	.000
Final	Parental education	-.324	.150	4.639	.031
	Growth on inflectional morphology	-.238	.087	7.558	.006

Note. There was missing data for 11 children. The regression analysis was therefore run with 109 participants.

Discussion and conclusion

Children with DLD experience language difficulties in the absence of a clear cause, but it is widely acknowledged that there is large heterogeneity within the group of children with DLD with respect to the severity of these difficulties, the affected language domains, and weaknesses beyond language. Far less is known about the variation in specialist educational support that children with DLD receive for their difficulties and factors which predict this variation. Using data from a longitudinal study including a cohort of monolingual and bilingual children with and without DLD between 5 and 8 years old, the current study set out to fill this gap.

We first aimed to establish for how many children with DLD in the cohort specialist educational provision was discontinued at age 7 or 8 years old. All children received specialist support when they were 5 or 6 years old, at wave 1 of the study, but this changed for a quarter of the sample two years later. Within two years, a large proportion of the children with DLD did not have specialist educational provision anymore. This contrasts with findings from the study by Dockrell et al. (2019) in the United Kingdom, who observed no differences in amount of support provided by schools over a two-year time period. Their study included children who were, on average, 10 years old at the first time point and who were thus substantially older than the participants of the current study. It may be that changes in specialist educational support more often occur at younger ages, when development is more dynamic and when the needs of a child have not yet crystallized. The severity of DLD and criteria for specialist educational provision may, however, also explain the differences in results between the current study and the study by Dockrell et al. (2019). The current study focused on a group of children with relatively severe difficulties, as children are otherwise not eligible for specialist educational support in the Netherlands. Such stringent criteria may result in more frequent changes in eligibility for support. However, alternatively, stringent criteria may also result in fewer changes in eligibility for support, because children who have more severe and persistent problems will be singled out and are less likely to improve.

Comparisons between monolingual and bilingual children with DLD showed that specialist educational provision was discontinued for nearly 30% of the monolingual children, while this was only 12.5% for the bilingual children. Although this difference fell just short of significance, this finding seems to indicate that bilingual children are more likely to receive specialist educational support for a longer period than monolingual children. This is interesting in view of the known overrepresentation of bilingual children in special education (Smeets et al., 2009). Comparisons in the full sample between children with DLD with and without specialist educational provision at wave 3 furthermore showed significant differences in parental education between the groups, with parents from children without provision having a higher level of education than parents from children with provision. Other differences in background characteristics were not significant. The effect of bilingualism and parental education both seem to indicate that environmental factors play a role in the discontinuation of specialist educational support for children with DLD. We will elaborate on this in our discussion of the third aim of this study.

The second aim was to investigate how children with DLD for whom specialist educational provision was discontinued perform on standardized and norm-referenced language and literacy measures, relative to TD children and children with DLD who did continue to receive specialist educational support. Our results showed clear differences between the two groups of children with DLD on the one hand and TD children on the other hand, both in the monolingual and bilingual group, with larger proportions of children with DLD scoring poorly on the standardized language and literacy measures than TD children. Except for the receptive vocabulary measure, the children with DLD with specialist educational support more often scored poorly than the children with DLD without this support. The difference was particularly striking on the word formation task assessing expressive inflectional morphology. These findings seem to reflect that severity of the language difficulties is considered when decisions on specialist educational provision are made (Dockrell et al., 2006; Stichting Siméa, 2017). They also suggest that, while their language difficulties may be less severe, the children with DLD in this sample without specialist educational provision cannot be classified as having 'transient' or 'resolved' DLD (Bishop & Edmundson, 1987). Both language and reading difficulties are observed in many children in this group, while they do not receive specialist educational support. Although it may be that they still do receive speech and language therapy or remedial teaching, this shows that this group of children remains vulnerable and requires attention. It may also be that these children function well without specialist provision, despite their lower test scores. For example, knowledge of language structures and rules, measured by such tests, may not automatically capture a child's communicative participation outcomes (Singer, Klatte, Welbie, Cnossen, & Gerrits, 2020). We were not able to include outcomes of communicative participation in the current study, while this may have played an important role in the decision process regarding provision.

Finally, the third aim of the study explored which linguistic, cognitive, and environmental factors predict the (dis)continuation of specialist educational provision of children with DLD. Parental education and scores on the word formation task, both at baseline and growth, turned out to be significant. The importance of the word formation task may, as previously mentioned, signal that children with more severe language difficulties are also more likely to have specialist educational provision during a longer period. Given the fact that early language abilities are highly predictive of later outcomes (McKean et al., 2017b), this suggests that those who need most support also receive most support. However, other linguistic variables, including sentence repetition, vocabulary, and narrative tasks were not found to be significant predictors in our model, while they are reported to be crucial for academic performance (Bishop & Edmundson, 1987; Dickinson & Tabors, 2002; West et al., 2021). It may be that the word formation measure is a proxy for a more complex or multi-factorial profile of difficulties. Additionally, there are several more specific reasons why the word formation task may be particularly sensitive to changes in specialist educational provision. First, word formation could be a better index of DLD severity than the other language measures, because the correct use of inflectional morphology is a core problem of children with DLD learning Dutch (de Jong, 1999). Inflectional morphology has, moreover, a rule-like nature, which shows a steep increase in development reflecting detection of the pattern and mastery of the rule (Rice, 2012). Children who struggle most in this area and do not show such growth may be the ones with the most persistent problems. In line with this idea, it could also be that difficulties in the area of morphology are most salient to professionals, increasing the likelihood that children making many morphological errors are evaluated as severe cases of DLD. Finally, we also consider the possibility that the word formation task taps into other underlying domain-general learning skills, such as procedural learning (e.g., Hamrick, Lum, & Ullman, 2018), which are relevant for the assessment of eligibility of specialist educational support.

Cognitive factors, including sustained attention and nonverbal intellectual functioning, were not found to be predictive of specialist educational support. This is not in line with our expectations, as learning capacity and nonverbal abilities are considered when decisions on specialist educational provision are made (Dockrell et al., 2006; Stichting Siméa, 2017). Moreover, Dockrell et al. (2019) also found nonverbal intellectual functioning to be a significant predictor of the number of hours of school support in the United Kingdom. Our findings, however, do not indicate that children with DLD without specialist educational provision benefit from more cognitive strengths, which could have compensated their linguistic weaknesses, than peers with provision. In contrast, parental education level was higher for children in this group, relative to children with DLD with educational provision. Given the importance of parental education for later language outcomes (Fisher, 2017), high parental education may be seen as a protective factor, increasing the likelihood of discontinuation of specialist educational support. Note that variables that predict the discontinuation of specialist educational provision can be labelled as protective factors if the children indeed do not need continuing support. It is also possible that children would still profit from additional support, even though they are not eligible anymore, and in this case, predicting variables (such as parental education) are risk factors. Another explanation for why parental education significantly predicted specialist educational provision is that decisions on (dis)continuation of support are made in agreement with parents (see also Dockrell et al., 2006) and that parents with a relatively high level of education are more likely to opt for regular education without specialist support for their child. Finally, higher levels of parental education may also indirectly reflect genetically determined compensatory abilities, such as strong planning skills, which could explain discontinuation of provision, but which were not included in the present study. Bilingualism was not found to be a significant predictor in our regression model. However, the difference in the proportion of monolingual and bilingual children with DLD for whom specialist educational provision was discontinued fell just short of significance and may indicate that professionals take amount of exposure in the school language into account when deciding on continuation of provision.

Our data show that specialist educational provision is terminated within a relatively short period of time for a significant proportion of children with DLD. We do not know whether the percentage of discontinuations we observe in the current study can be generalized to the population of children with DLD attending special education, but given the size of our sample, we believe our figures are certainly indicative. The language and literacy scores show that the children for whom provision was discontinued cannot be qualified as children with a 'resolved DLD'. One question that can be asked is how the further educational career of these children progresses, and what their final levels are. To our knowledge, no research has yet been conducted in the Netherlands into the long-term outcomes regarding education (and also professional practice and social success) in children with DLD. Such research is desirable, especially when it makes a distinction between those individuals who have participated in special education for a longer period, in comparison to those who were referred to lighter or no educational arrangements.

It is interesting and important to understand the grounds on which professionals decide whether a child can do without specialist educational provision. Our results suggest that mono- versus bilingualism plays a role, as does parental education level. In the Dutch context, these two variables are not easy to separate, so further research is recommended. It is remarkable that a child's performance in the morphological domain, and progress therein, is predictive of the decision to discontinue specialist educational support, while other linguistic and cognitive factors were not. Of course, we do not know whether the predictors that our analysis identified as significant correspond to the factors professionals consider when deciding whether a child can be discharged from special education. In a follow-up study, it therefore seems useful to us to have professionals in special education reflect on the decision-making process regarding (dis)continuation of specialist support, as has been done in the United Kingdom (Dockrell et al., 2006). The question is whether what they see as relevant factors, related to characteristics of the child itself, their family and socio-cultural background, are consistent with the results reported here. Such a study could contribute to strengthening the evidence base for decision-making regarding the guidance of children with DLD.

Author contributions

Conceptualization, E.B.; Formal analysis, T.B.; Funding acquisition, E.B.; Investigation, T.B.; Methodology, T.B. and E.B.; Supervision, E.B. and F.W.; Writing - original draft, T.B. and E.B.; Writing - review & editing, T.B., F.W., and E.B.

Literature

Aljahlan, Y., & Spaulding, T. J. (2021). Attentional shifting in children with developmental language disorder: A meta-analysis. *Journal of Communication Disorders, 91*, 106105. doi.org/10.1016/j.jcomdis.2021.106105

Beitchman, J. H., Wilson, B., Brownlie, E. B., Walters, H., Inglis, A., & Lancee, W. (1996). Long-term consistency in speech/language profiles: II. Behavioral, emotional, and social outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*(6), 815-825. doi.org/10.1097/00004583-199606000-00022

Bishop, D. V. (2010). Overlaps between autism and language impairment: phenomimicry or shared etiology?. *Behavior genetics, 40*(5), 618-629. doi.org/10.1007/s10519-010-9381-x

Bishop, D. V. (2009). Genes, cognition, and communication: insights from neurodevelopmental disorders. *Annals of the New York Academy of Sciences, 1156*(1), 1-18. https://www.doi.org//10.1111/j.1749-6632.2009.04419.x

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of speech and hearing disorders, 52*(2), 156-173. doi.org/10.1044/jshd.5202.156

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Catalise-2 Consortium, Adams, C., ... & house, A. (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*(10), 1068-1080. doi.org/10.1111/jcpp.12721

Blom, E., & Boerma, T. (2020). Do children with Developmental Language Disorder (DLD) have difficulties with interference control, visuospatial working memory, and selective attention? Developmental patterns and the role of severity and persistence of DLD. *Journal of Speech, Language, and Hearing Research, 63*(9), 3036-3050. doi.org/10.1044/2020_JSLHR-20-00012

Blom, E., Boerma, T., & de Jong, J. (2020). Multilingual Assessment Instrument for Narratives (MAIN) adapted for use in Dutch. *ZAS Papers in Linguistics, 64*, 51-56. doi.org/10.21248/zaspil.64.2020.557

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research, 58*(6), 1747-1760. doi.org/10.1044/2015_JSLHR-L-15-0058

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Impli-cations for clinical practice. *International journal of language & communication disorders, 51*(6), 626-638. doi.org/10.1111/1460-6984.12234

Boerma, T., Leseman, P., Wijnen, F., & Blom, E. (2017). Language proficiency and sustained attention in monolingual and bilingual children with and without language impairment. *Frontiers in Psychology, 8*, 1241. doi.org/10.3389/fpsyg.2017.01241

Bornstein, M. H., Hahn, C. S., Putnick, D. L., & Suwalsky, J. T. (2014). Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child development, 85*(4), 1346-1356. doi.org/10.1111/cdev.12192

Botting, N. (2005). Non-verbal cognitive development and language impairment. *Journal of child psychology and psychiatry, 46*(3), 317-326. doi.org/10.1111/j.1469-7610.2004.00355.x

Botting, N. (2020). Language, literacy and cognitive skills of young adults with developmental language disorder (DLD). *International journal of language & communication disorders, 55*(2), 255-265. doi.org/10.1111/1460-6984.12518

Brus, B. T., & Voeten, M. J. M. (1999). *Een-minuut-test: vorm A en B (EMT)*. Nijmegen, The Netherlands: Berkhout.

Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairments. doi.org/10.1044/1092-4388(2008/07-0259)

Chiat, S. (2015). Nonword repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp.125–150). Bristol, England: Multilingual Matters. doi.org/10.21832/9781783093137

Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders–a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of child psychology and psychiatry, 46*(2), 128-149. doi.org/10.1111/j.1469-7610.2004.00342.x

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of child psychology and psychiatry, 42*(6), 741-748. doi.org/10.1111/1469-7610.00770

Conti-Ramsden, G., Botting, N., Simkin, Z., & Knox, E. (2001). Follow-up of children attending infant language units: Outcomes at 11 years of age. *International Journal of Language & Communication Disorders, 36*(2), 207-219. doi.org/10.1080/13682820121213

Curtis, P. R., Frey, J. R., Watson, C. D., Hampton, L. H., & Roberts, M. Y. (2018). Language disorders and problem behaviors: A meta-analysis. *Pediatrics, 142*(2). doi.org/10.1542/peds.2017-3551

Dickinson, D. K., & Tabors, P. O. (2002). Fostering language and literacy in classrooms and homes. *Young Children, 57*(2), 10-19.

Dockrell, J. E., & Lindsay, G. (2008). Inclusion versus specialist provision for children with developmental language disorders. In: Norbury, C. F., Tomblin, J. B., & Bishop, D. V. M. (eds). *Understanding developmental language disorders. From theory to practice,* 131-147. London: Psychology Press.

Dockrell, J. E., Lindsay, G., Letchford, B., & Mackie, C. (2006). Educational provision for children with specific speech and language difficulties: perspectives of speech and language therapy service managers. *International Journal of Language & Communication Disorders, 41*(4), 423-440.

Dockrell, J. E., Ricketts, J., Palikara, O., Charman, T., & Lindsay, G. A. (2019). What drives educational support for children with developmental language disorder or autism spectrum disorder: Needs, or diagnostic category?. *Frontiers in Education, 4,* 29. doi.org/10.3389/feduc.2019.00029

Dubois, P., St-Pierre, M. C., Desmarais, C., & Guay, F. (2020). Young adults with developmental language disorder: a systematic review of education, employment, and independent living outcomes. *Journal of Speech, Language, and Hearing Research, 63*(11), 3786-3800. doi.org/10.1044/2020_JSLHR-20-00127

Duinmeijer, I., de Jong, J., & Scheper, A. (2012). Narrative abilities, memory and attention in children with a specific language impairment. *International Journal of Language & Communication Disorders, 47*(5), 542-555. doi.org/10.1111/j.1460-6984.2012.00164.x

Eadie, P., Conway, L., Hallenstein, B., Mensah, F., McKean, C., & Reilly, S. (2018). Quality of life in children with developmental language disorder. *International journal of language & communication disorders, 53*(4), 799-810. doi.org/10.1111/1460-6984.12385

Ebbels, S. H., McCartney, E., Slonims, V., Dockrell, J. E., & Norbury, C. F. (2019). Evidence-based pathways to intervention for children with language disorders. *International journal of language & communication disorders, 54*(1), 3-19.

Ebert, K. D., & Kohnert, K. (2011). Sustained attention in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 54*(5), 1372–1384. doi.org/10.1044/1092-4388(2011/10-0231)

Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: a meta-analysis. *Journal of Speech, Language, and Hearing Research, 50*(1), 177-195. doi.org/1092-4388/07/5001-0177

Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *Journal of Speech, Language, and Hearing Research, 60*(10), 2935–2948. doi.org/10.1044/2017_JSLHR-L-16-0310

Gagarina, N. V., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., ... & Walters, J. (2012). MAIN: Multilingual assessment instrument for narratives. *ZAS papers in linguistics, 56*, 155-155. doi.org/10.21248/zaspil.56.2019.414

Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 57*(4), 1363-1382. doi.org/10.1044/2014_JSLHR-L-12-0363

Gerrits, E., de Jong, J., Zwitserlood, R., & Klatte, I. (2019). The Netherlands. In: Law, J., McKean, C., Murphy, C.-A., & Thordardottir, E. (Eds), *Managing children with developmental language disorder: Theory and practice across Europe and beyond* (1st ed.), Routledge, London, pp. 339-350.

Gilger, J. W., & Kaplan, B. J. (2001). Atypical brain development: a conceptual framework for understanding developmental learning disabilities. *Developmental neuropsychology, 20*(2), 465-481. doi.org/10.1207/S15326942DN2002_2

Hamrick, P., Lum, J. A., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences, 115*(7), 1487-1492.

IBM Corp. (2016). *IBM SPSS Statistics for Windows.* Version 24.0. Armonk: IBM Corp.

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., ... & Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research, 42*(3), 744-760.

Law, J., McKean, C., Murphy, C.-A., & Thordardottir, E. (2019). *Managing children with developmental language disorder: Theory and practice across Europe and beyond* (1st ed.), Routledge, London.

Leonard, L. B. (2014). *Children with specific language impairment.* Cambridge, Mass.: MIT press.

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the “specifics” of specific reading disability and specific language impairment. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(7), 869-874. doi.org/10.1111/1469-7610.00674

McKean, C., Wraith, D., Eadie, P., Cook, F., Mensah, F., & Reilly, S. (2017a). Subgroups in language trajectories from 4 to 11 years: The nature and predictors of stable, improving and decreasing language trajectory groups. *Journal of Child Psychology and Psychiatry, 58*(10), 1081-1091. doi.org/10.1111/jcpp.12790

McKean, C., Reilly, S., Bavin, E. L., Bretherton, L., Cini, E., Conway, L., ... & Mensah, F. (2017b). Language Outcomes at 7 Years: Early Predictors and Co-Occurring Difficulties. *Pediatrics, 139*(3), e20161684. doi.10.1542/peds.2016-1684

Mueller, K. L., & Tomblin, J. B. (2012). Examining the comorbidity of language disorders and ADHD. *Topics in Language Disorders, 32*(3), 228.

doi.org/10.1097/TLD.0b013e318262010d

Nederlands Jeugdinstituut (NJI) (2022). *Cijfers over speciaal onderwijs: cluster 2.* Retrieved from: https://www.nji.nl/cijfers/speciaal-onderwijs-cluster-2.

Norbury, C. F., Vamvakas, G., Gooch, D., Baird, G., Charman, T., Simonoff, E., & Pickles, A. (2017). Language growth in children with heterogeneous language disorders: a population study. *Journal of Child Psychology and Psychiatry, 58*(10), 1092-1105. doi.org/10.1111/jcpp.12793

Pauls, L. J., & Archibald, L. M. (2016). Executive functions in children with specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 59*(5), 1074-1086. doi.org/10.1044/2016_JSLHR-L-15-0174

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research, 39*(6), 1239-1257. doi.org/10.1044/jshr.3906.1239

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2; 6 to 21 years of age. *Journal of Speech, Language, and Hearing Research, 58*(2), 345-359. doi.org/10.1044/2015_JSLHR-L-14-0150

Scarborough, H. S., & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech, Language, and Hearing Research, 33*(1), 70-83. doi.org/10.1044/jshr.3301.70

Schlichting, L. (2005). *Peabody Picture Vocabulary Test—III NL.* Amsterdam, the Netherlands: Harcourt Test Publishers.

Singer, I., Klatte, I. S., Welbie, M., Cnossen, I. C., & Gerrits, E. (2020). A multidisciplinary Delphi consensus study of communicative participation in young children with language disorders. *Journal of Speech, Language, and Hearing Research, 63*(6), 1793-1806. doi.org/10.1044/2020_JSLHR-19-00326

Smeets, E., Driessen, G., Elfering, S. & Hovius, M. (2009). *Allochtone leerlingen en speciale onderwijsvoorzieningen.* Nijmegen: ITS.

Snowling, M. J., Bishop, D. V. M., Stothard, S. E., Chipchase, B., & Kaplan, C. (2006). Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *Journal of Child Psychology and Psychiatry, 47*(8), 759-765. doi.org/10.1111/j.1469-7610.2006.01631.x

Snowling, M. J., Duff, F. J., Nash, H. M., & Hulme, C. (2016). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry, 57*(12), 1360-1369. doi.org/10.1111/jcpp.12497

Stichting Siméa (2014). *Van vraag naar ondersteuning. Landelijk kader inrichting passend onderwijs.* https://simea.nl/media/richtlijnen/simea-brochure-van-vraag-naar-ondersteuning-nov-2014_.pdf

Stichting Siméa (2016). *Handreiking meertaligheid TOS.* Retrieved from: http://www.simea.nl/dossiers/si166289-simea-handreiking-meertaligheid-tos.pdf

Stichting Siméa (2017). *Richtlijn toelaatbaarheid.* https://simea.nl/media/richtlijnen/simea-brochure-richtlijn-toelaatbaarheid-20170901.pdf

Stothard, S. E., Snowling, M. J., Bishop, D. V., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41*(2), 407-418. doi.org/10.1044/jslhr.4102.407

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., & Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research, 64*(9), 3578-3593. doi.org/10.1044/2021_JSLHR-20-00552

Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The Stability of Primary Language Disorder: Four Years After Kindergarten Diagnosis. *Journal of Speech, Language, and Hearing Research, 46*(6), 1283-1296. doi.org/10.1044/1092-4388(2003/100)

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.), *Methods for assessing multilingual children: disentangling bilingualism from Language Impairment*, (pp. 301-30). Bristol: Multilingual Matters.

Van den Bos, K.P., Lutje Spelberg, H.C., Scheepstra, A. J. M., & De Vries, J.R. (1994). *De Klepel, Vorm A en B. Een test voor de leesvaardigheid van pseudowoorden. Verantwoording, Handleiding, Diagnostiek en Behandeling.* Amsterdam: Pearson.

Verhoeven, L., & Vermeer, A. (2001). *Taaltoets Alle Kinderen* (TAK). Arnhem: Cito.

Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities, 34*(9), 2586-2597. doi.org/10.1016/j.ridd.2013.05.014

Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research, 42*(3), 744-760. doi.org/10.1044/jslhr.4203.744

Webster, R. I., Erdos, C., Evans, K., Majnemer, A., Kehayia, E., Thordardottir, E., ... & Shevell, M. I. (2006). The clinical spectrum of developmental language impairment in school-aged children: language, cognitive, and motor findings. *Pediatrics, 118*(5), e1541-e1549. doi.org/10.1542/peds.2005-2761

Wechsler, D., & Naglieri, J.A. (2008). *Wechsler Nonverbal Scale of Ability.* Purchased from: http://www.pearsonclinical.nl/wnv-nl-wechsler-non-verbal

West, G., Shanks, D. R., & Hulme, C. (2021). Sustained attention, not procedural learning, is a predictor of reading, language and arithmetic skills in children. *Scientific Studies of Reading, 25*(1), 47-63. doi.org/10.1080/10888438.2020.1750618

Wood, P., & Bates, S. (2020). National and international approaches to special education needs and disability provision. *Education 3-13, 48*(3), 255-257. doi.org/10.1080/03004279.2019.1664395

Zambrana, I. M., Pons, F., Eadie, P., & Ystrom, E. (2014). Trajectories of language delay from age 3 to 5: Persistence, recovery and late onset. *International Journal of Language & Communication Disorders, 49*(3), 304-316. doi.org/10.1111/1460-6984.12073

Stem-, Spraak- en Taalpathologie

Vol. 27, 2022, Special Issue, pp. 24-52
Special Issue for Steven Gillis

Gepubliceerd online: 5 november 2022

University of Groningen Press
https://doi.org/10.21827/32.8310/2022-SG-24

Word-level spelling of children with dyslexia and developmental language disorder

Elise de Bree[1], Imme Lammertink[2,3],
Merel van Witteloostuijn[1], & Judith Rispens[3]

[1] *Department of Education and Pedagogy, Utrecht University*
[2] *Royal Dutch Kentalis, Utrecht*
[3] *Faculty of Humanities, University of Amsterdam*

Samenvatting

Veel kinderen met dyslexie en met een taalontwikkelingsstoornis (TOS) hebben moeite met spellen. Het is niet bekend of deze problemen hetzelfde zijn voor beide groepen, omdat er weinig directe spellingvergelijkingen tussen deze twee groepen kinderen zijn gemaakt. In deze studie vergeleken we de woorddictee-uitkomsten en foutenpatronen van kinderen (leeftijdsrange 8.2-10.4 jaar) met dyslexie (n=31), TOS (n=30) en een leeftijdsgematchte controlegroep (n=31). Daarnaast verdeelden we de TOS-groep in kinderen met TOS met leesproblemen (TOS+LP, n=18) en zonder leesproblemen (alleen-TOS, n=12). Tot slot voerden we regressieanalyses uit om te bepalen welke taal- en lees(gerelateerde) vaardigheden invloed hebben op spellingsuitkomsten.

Zowel de dyslexie- als TOS-groep vertoonden spellingproblemen. Deze waren het meest uitgesproken voor de TOS+LP-groep. Wat betreft foutsoorten maakten de dyslexie- en TOS-groepen meer fouten in alle categorieën dan de controlegroep. De dyslexie, alleen-TOS en TOS+LP-groepen lieten eenzelfde patroon zien wat betreft het spellen van woorden met foneem-grafeemassociaties die niet helemaal klankzuiver zijn (gauw als *gouw), met woorden waarin fonologie-orthografie-en morfologie moeten worden gecombineerd (pittig als *pitteg) en met regelwoorden (metro als *meetro). De dyslexie en TOS+LP-groepen lieten ook meer fouten zien in klankzuivere foneemgrafeemkoppelingen (boek als *beok) en met orthografische kennis (leenwoorden). De regressieanalyses lieten een sterke bijdrage zien van woordlezen en snelbenoemen aan woordspelling. Regressieanalyses zonder deze lees(gerelateerde) vaardigheden lieten een beperkte bijdrage van morfologie zien aan de spellingsuitkomsten.

Deze resultaten bevestigen dat kinderen met dyslexie en TOS spellingprobemen hebben, zeker kinderen met TOS+LP. Ze wijzen naar de belangrijke bijdrage van woordleesvaardigheid en snelbenoemen voor spellen, en een beperktere bijdrage van morfologie.

Correspondentieadres:
Elise de Bree
Utrecht University
Department of Education and Pedagogy
P.O. Box 80140, 3580TC Utrecht, The Netherlands
E-mail: e.h.debree@uu.nl

Samen geven ze de noodzaak weer van systematisch spellingonderwijs en ondersteuning voor kinderen met dyslexie en TOS, alsook het belang van mondelinge taalondersteuning.
Trefwoorden: Spelling, Dyslexie, Taalontwikkelingsstoornis, Lezen, Taal

Abstract

Spelling is challenging for children with dyslexia and for children with Developmental Language Disorder (DLD), but it is not clear as yet whether the difficulties are the same in both groups of children, as few direct comparisons between spelling outcomes of these groups have been reported. We therefore compared the mean word dictation spelling outcomes of children (age range 8.2-10.4 years) with dyslexia (n=31), DLD (n=30) and typically developing (TD) age-matched peers (n=31) as well as the error types. Additionally, we divided the group of children with DLD into those with (DLD+RD, n=18) and without word reading difficulties (DLD-only, n=12). Finally, we conducted concurrent regression analyses to establish which language- and literacy-related abilities contribute to spelling.

Both the groups with dyslexia and DLD displayed spelling problems. These were most pronounced for the subgroup of children with DLD+RD. The error type analyses showed that the groups with dyslexia and DLD made more errors in all categories than the TD group. Furthermore, the dyslexia, DLD-only and DLD+RD groups showed similar errors with phoneme-grapheme associations that were not entirely transparent (gauw [quick] as *gouw), with mappings of phonology-orthography and morphology (pittig [spicy] as *pitteg), and errors concerning spelling rules (metro [metro] as *meetro). The dyslexia and DLD+RD groups also showed more errors in transparent phonology-grapheme mappings (boek [book] as *beok) and orthographic knowledge (loanwords, hyphenation). The regression analysis showed a strong contribution of word reading and rapid automatized naming to spelling. When these measures were excluded, morphology contributed significantly, but to a lesser extent.

The findings confirm the spelling problems of children with dyslexia and with DLD, especially the severe problems of children with DLD+RD. They also point to the importance of word-reading ability and rapid automatized naming for spelling, as well as to a smaller contribution of morphology. Together, the findings show the need for systematic spelling instruction and intervention for children with dyslexia and DLD, as well as support of oral language.
Keywords: Spelling, Dyslexia, Developmental Language Disorder, Reading, Language

Introduction

Being able to write is an important component of a learner's language ability. For the writer, more effortful spelling adds to the cognitive load during the writing process. This, in turn, might affect text production fluency and quality, leading to written texts that are more difficult to understand (Berninger & Swanson, 1994; Graham et al., 1997). For the reader, correct spelling ensures focus on the content of a message; spelling errors disturb this process (Graham et al., 2011). Learning to spell is therefore essential and requires instruction (Graham & Santangelo, 2014). However, despite instruction, not everyone becomes a fluent speller, especially when spelling words that are not spelled like they sound. In this study, we compare spelling performance of two groups of children who are likely to exhibit spelling difficulties, children with dyslexia and children with developmental language disorder (DLD), to that of typically developing (TD) peers.

Spelling acquisition

Models of spelling (and reading) development state the importance of phonology (phonemes) and orthography (graphemes) and their associations (Ehri, 2000; Perfetti & Hart, 2002; Seidenberg & McClelland, 1989; Treiman & Kessler, 2014). Indeed, phonological skills, rapid automatized naming (RAN) and word reading have been reported to be important contributors to spelling outcomes (Georgiou et al., 2020; Lervåg & Hulme, 2010; Moll et al., 2014). Furthermore, interventions focused on phonological skills, phoneme-grapheme connections and word reading generally lead to improvement of spelling (meta-analyses by Galuschka et al., 2014, 2020). These findings indicate that word reading and reading-related cognitive variables contribute to spelling.

Oral language abilities are also assumed to contribute to spelling outcomes, as connections between spoken/oral language and orthography are assumed to facilitate orthographic knowledge (Perfetti & Hart, 2002; Seidenberg & McClelland, 1989; Treiman & Kessler, 2014). Studies have confirmed that spelling outcomes are influenced by morphosyntax and vocabulary (Apel et al., 2012; Ouellette, 2010; van Weerdenburg et al., 2011) and that interventions in morphology positively affect spelling outcomes (Bowers et al., 2010; Devonshire & Fluck, 2010; Goodwin & Ahn, 2013; Galuschka et al., 2020). Nevertheless, direct effects of morphosyntactic abilities and vocabulary on spelling outcomes are not always attested, especially when other factors, such as orthographic knowledge, are accounted for (Kim et al., 2013; McArthy et al., Werfel et al., 2021).

Spelling in developmental dyslexia

One group of children who show spelling difficulties are children with dyslexia, referring to children who show severe and persistent difficulties with word reading and spelling despite adequate literacy instruction (American Psychiatric Association, 2013; Lyon et al., 2003). As the primary difficulty of children with developmental dyslexia resides in grapheme-phoneme associations, measures tapping these associations (word reading and RAN) can be taken as cognitive correlates of spelling difficulties in this group.

The relationship between oral language abilities and spelling in children with dyslexia is less clear. One language domain at risk for suboptimal development in dyslexia is vocabulary (Duff et al., 2015; Torppa et al., 2010; van Viersen et al., 2017, but see a recent study showing that vocabulary deficits are not necessarily attested in children with dyslexia (Adlof et al., 2021)). Similarly, for morphology, some studies have reported that children with dyslexia have more difficulty with morphology than their age-matched peers without dyslexia (Casalis et al., 2004; Fowler & Liberman, 1995), but these difficulties are not always present or severe and persistent in dyslexia (Elbro & Arnbak, 1996; Quémart & Casalis, 2013) or are only present in those children with dyslexia who have comorbid language difficulties (Catts et al., 2005). It thus seems that for children with dyslexia, orthographic difficulties are most pervasive, and the role of suboptimal development of oral language as a potential source for spelling problems is less clear.

Different spelling error types have been reported for children with dyslexia. For instance, phonological errors have been found (Caravolas & Volín, 2001; Hoefflin & Franck, 2005)[1], but these seem to occur mostly in more opaque orthographies. Morphosyntactic and orthographic errors[2] are more frequent, especially in more transparent orthographies (Angelelli et al., 2004; Landerl & Wimmer, 2000; Protopapas et al., 2013). It has been proposed that the difficulties of children with dyslexia indicate a delay in internalizing the systematicity of the orthographic system (Protopapas et al., 2013). This refers to spelling words that are sensitive to the orthographic context and to words that require application of orthographic rules (also on the basis of morphological inflection). Thus, the spelling difficulties that children with dyslexia display have a clear literacy basis but may also be influenced by oral language abilities.

Spelling in Developmental Language Disorder

A second group of children who show spelling difficulties are children with DLD. These children are diagnosed with difficulties in multiple areas of oral language, including vocabulary and morphosyntax in the absence of a known biomedical cause, intellectual disability, and/or unfavourable psycho-social/educational conditions (Bishop et al., 2017). The term DLD is now commonplace.[3] Two recent meta-analyses show that children with DLD often show lower spelling outcomes than their age-matched peers (Joye et al., 2019; Graham et al., 2020). As a substantial number of children with DLD also develop reading difficulties (Eisenmajer et al., 2005; McArthur et al., 2000; Snowling et al., 2019; cf. Catts et al., 2005), it is likely that part of these spelling difficulties stem from literacy-related difficulties. Indeed, Vandewalle and colleagues (2011) found that Flemish children with both language and reading disorders obtained lower spelling outcomes than children with a language disorder only. Furthermore, the meta-analysis by Joye et al. (2019) confirmed the tight relationship between reading and spelling difficulties in children with DLD as well as the role that phonology plays in spelling.

Findings on the extent to which language difficulties of children with DLD influence their

[1]Phonological errors are errors in which the word's phonological form is altered, leading the written word to be pronounced differently from the one intended (i.e., phonetically incorrect).

[2]Orthographic errors refer to words in which the word's correct pronunciation is maintained, but in which the written representation is altered by substituting alternative graphemes for the same phonemes (i.e., phonetically accurate).

[3]We will refer to DLD when describing previously conducted studies, even when previous studies have referred to these children as children with specific language impairment.

spelling outcomes are less pronounced. There are indications that oral language abilities contribute to spelling. For instance, van Weerdenburg and colleagues (2011) showed that several language abilities (verbal-sequential processing, auditory perception, speech production and lexical-semantic abilities) contributed to spelling outcomes seven months later for 7-year-old children with DLD and controls. It should, however, be noted that in their study, verbal-sequential processing had the highest predictive value of spelling outcome. This factor contained RAN, a task strongly related to word reading (Moll et al., 2014). Also, in the study by van Weerdenburg et al. (2011) word reading was not entered as a contributor of spelling, so it cannot be assessed whether both language and literacy contributed to spelling. Furthermore, a study by Larkin et al. (2013) did not find a contribution of vocabulary and morphosyntax to concurrent spelling outcomes in their sample of 11-year-old children with DLD and two control groups. Similarly, in a study by Werfel et al. (2021), the spelling outcomes of first-grade children with DLD were determined mainly by word reading and phonological awareness, not vocabulary and morphosyntax. These findings are therefore not conclusive about the contribution of language abilities to spelling outcomes in children with DLD.

With respect to the types of errors children with DLD make in their spelling, phonologically unacceptable errors seem to be more prevalent than orthographically unacceptable errors (Broc et al., 2013; Joye et al., 2020; Larkin & Snowling, 2008; McCarthy et al., 2012). Also, children with DLD have difficulty in spelling morphologically inflected targets (Deacon et al., 2014; Joye et al., 2020; Larkin et al., 2013; Silliman et al., 2006; Windsor et al., 2000). These findings on spelling errors have been confirmed in a recent scoping review (Broc et al., 2021). Thus, children with DLD's language difficulties are visible in both oral and written language (Windsor et al., 2000), but it is not clear whether both oral language and literacy skills (Bishop & Clarkson, 2003; Silliman et al., 2006) or mainly literacy skills (Mackie et al., 2013; McCarthy et al., 2012; Werfel et al., 2021) contribute to their spelling outcomes.

Comparing spelling in children with dyslexia and children with DLD

Spelling difficulties are likely to occur both for children with dyslexia and for children with DLD. One question is whether the difficulties are (partly) the same. Relatively few direct comparisons between spelling outcomes of these groups have been reported. Such information is needed to inform spelling instruction and additional intervention, as it needs to be clear whether the areas of difficulty (and thus required instruction) are the same or different. Some studies have compared teenagers and adolescents' spelling (Goulandris et al., 2007; Puranik et al., 2007) and most concern children at primary school age (Alloway, 2017[4]; Bishop et al., 2009; Larkin & Snowling, 2008; McCarthy et al., 2012; Scuccimarra et al., 2008; Snowling et al., 2019). The comparisons have rendered mixed findings. One pattern of findings is that the DLD group performs as poorly on spelling as the group with dyslexia

[4]Alloway et al. (2017) do not provide statistical testing. Calculation of effect sizes between the groups show that their sample of children with DLD performed more poorly than their sample of children with dyslexia (d'= .487) and their control group (d'= -2.249). The spelling outcomes of children with dyslexia were also lower than those of the control group (d'= -1.427), but the effect size was smaller than that of DLD and control group.

(Goulandris et al., 2007; Larkin & Snowling, 2008; Puranik et al., 2007), another that the DLD group shows poorer spelling than the control group, but still better than that of children with dyslexia (Snowling et al., 2019) and a third that the DLD group outperforms the dyslexic groups and cannot be distinguished from the group of typically developing peers (Bishop et al., 2009; McCarthy et al., 2012).

Furthermore, studies that have included both groups with one disorder (dyslexia-only and DLD-only) as well as a combined group (DLD+poor word reading/dyslexia[5]) show a pattern of DLD-only obtaining better spelling results than dyslexia-only and DLD+poor word reading showing poorest performance (Bishop et al., 2009; McCarthy et al., 2012; Snowling et al., 2019). Furthermore, Brizzolara et al. (2011) and Scuccimarra et al. (2008) found that children with DLD+poor word reading and dyslexia-only obtained similar spelling outcomes, which were significantly lower than those of typically developing peers. Together, the findings suggest that children with combined DLD+poor word reading show poorest spelling performance, that children with dyslexia-only show spelling problems and that they are present for children with DLD-only, but less severe.

All previous spelling comparisons of dyslexia and DLD included comparisons of mean spelling outcomes between the groups. Only two studies compared the types of errors made (Larkin & Snowling, 2008; McCarthy et al., 2012). The findings of these two studies are not the same: Larkin and Snowling (2008) found that their DLD group obtained a lower score compared to the dyslexia and control group on the measure of phonologically acceptable spellings, possibly indicating a more severe phonological deficit in this group. In contrast, findings of McCarthy et al. (2012) showed that the DLD and TD groups did not differ from each other on most error types (phonological, graphemic, semantic error), except the orthographic pattern errors (DLD > TD).[6] Both groups outperformed the DLD+dyslexia and dyslexia groups on all error categories, except the graphemic errors, in which the pattern was reversed. The DLD+dyslexia and the dyslexia groups did not differ from each other. As findings are not consistent, further comparisons between spelling outcomes of children with dyslexia and DLD are needed to clarify whether (and how) the spelling difficulties overlap.

Present study

In the present study, we aimed to answer the question to what extent spelling outcomes are related to literacy and/or oral language difficulties. Three specific questions were addressed. The first was whether children with dyslexia, DLD and TD children show differences in mean spelling dictation outcomes. Although we expect lower outcomes for the DLD and dyslexia

[5]Studies vary in the inclusion criteria of children with poor word reading; some rely on poor word reading at one time point, whereas others rely on a formal diagnosis of dyslexia, in which severity of the word reading difficulties as well as persistence/didactic resistance are established. We use the general category of poor word reading and refer to dyslexia when this has been formally established.

[6]In their study, this category referred to incorrect sound-symbol correspondences (beb for bed), incorrect rules for combining letters (whent for went), incorrect patterns that govern spelling within the root/base word (lader for ladder) and incorrect positional constraints on spelling patterns (neckst for next) (McCarthy et al., 2012, p. 16).

groups compared to the spelling of TD children, we expect more heterogeneity in the DLD group (Joye et al., 2019) than in the group of children with dyslexia. A subgroup of children with DLD+poor word reading is expected to be present and is expected to perform most poorly on spelling (Bishop et al., 2009; McCarthy et al., 2012; Snowling et al., 2019).

The second question was whether the pattern of errors differed for the children with dyslexia and with DLD. On the basis of the literature, both the dyslexia and DLD groups are likely to show more difficulty with targets that are non-phonologically transparent. For the dyslexia group, errors might be especially pervasive for words requiring knowledge of (morphological-)orthographic rules and conventions (Protopapas et al., 2013). For the DLD group, errors are anticipated in words requiring inflection in general (Deacon et al., 2014; Joye et al., 2020), including those that use orthographic conventions for morphological inflections. Furthermore, words with irregular orthographic patterns might also be difficult for this group, as they demand connections between semantics and orthography (Wolter & Apel, 2010).

The final question was whether across all children spelling outcomes are predicted by literacy(-related) abilities as well as oral language abilities or whether there is a dominant contribution only of the literacy(-related) abilities. Word reading is expected to be a dominant contributor to spelling outcomes (Georgiou et al., 2020), as is the reading proxy of RAN (Moll et al., 2014). With respect to oral language abilities, there is some evidence for a unique contribution to spelling (van Weerdenburg et al., 2011), but this finding has not been consistent (Kim et al., 2013; Larkin et al., 2013; McCarthy et al., 2012; Werfel et al., 2021). We included measures of general cognitive ability as control variables: nonverbal intelligence, sustained attention and short-term and working memory in the verbal and visuospatial domain.

Methods

Participants

There were three groups of children: 1) a group of children with dyslexia ($n = 31$, 13 females, mean age 9.4 years, SD .58, range 8.4-10.4), 2) a group of children with DLD ($n = 30$, 8 females, mean age 9.3 years, SD .59, range 8.2-10.4), and a group of TD age-matched peers ($n = 31$, 10 females, mean age 9.3 years, SD .63, range 8.2-10.4). Children were included if at least one of their parents was a native speaker of Dutch and if Dutch was the dominant home language (next to Dutch-only at school). Children with dyslexia were recruited through treatment centres and Facebook support groups for parents. Children with DLD were recruited through four national DLD organizations in the Netherlands, an association for parents of children with DLD, and through self-employed speech therapists. The typically developing children were recruited from three different primary schools across the Netherlands.

The groups did not differ in age ($F(2, 91) = .344$, $p = .710$) and in gender distribution ($\chi(2) = 1.643$, $p = .442$). Socio-economic status (SES) was determined on the basis of the children's home or school postal codes, depending on the testing location. The reported SES scores indicate the social status of a given neighbourhood in comparison to other neighbourhoods in the Netherlands: the score consists of a combined score taking mean income, mean educational level and mean working status of the people living in a particular postal code into account (NISR, 2017, open source data). The mean score is 0, with higher scores indicating higher socio-economic status. Mean SES for the DLD group was .208 (.79), for the dyslexia group .247 (1.21) and for the TD group -.063 (1.05). There was no main effect of SES (F(2, 89) = .816, p = .446, $\eta_p^2 = .018$).

Children with dyslexia all had a prior formal diagnosis of dyslexia. Such a diagnosis is made by a licensed clinician (psychologist) on the basis of establishing both the severity (≤10th percentile) and the persistence of the literacy deficit, as well as on the basis of excluding other factors that might impact on opportunities to learn to read and spell (SDN et al., 2016). All children with dyslexia only had a diagnosis of dyslexia, meaning that there was no comorbid DLD (or other disorders). Three children with dyslexia spoke an additional language next to Dutch.

Children with DLD all had a prior formal diagnosis of DLD made by a licensed clinician. The criteria of diagnosis were (a) performance at least 1.5 standard deviations below the norm on two out of four subscales (speech production, auditory processing, grammatical knowledge, lexical semantic knowledge) of a standardized language assessment test battery administered by a licensed clinician, (b) no diagnosis of autism spectrum disorder, attention deficit hyperactivity disorder, or other physiological problems, c) nonverbal intelligence in the normal range (NVLF, 2016; Siméa, 2017). None of the children spoke another language next to Dutch. Two of the children with DLD had been formally diagnosed with dyslexia and for eight, there were documented suspicions of developmental dyslexia by the speech and language therapists.

The group of TD children was age-matched to the children with dyslexia and with DLD. It consisted of children without parental or teacher concerns regarding language and literacy development and no known behavioural problems. The children were randomly selected for participation in order to ensure that our sample included children across the range of language and literacy abilities.

We administered literacy, language, and non-verbal intelligence measures to confirm group inclusion and to compare the groups. With respect to reading, a timed word reading task (Eén Minuut Test, EMT, [One Minute Test]; Brus & Voeten, 1999) and timed pseudoword reading task (Klepel; van den Bos et al., 1994) were used. The cut-off score for poor performance was a standard score ≤ 6 (≤10th percentile), see Instruments and Table 1. A MANOVA with timed (pseudo-)word reading and group (Dyslexia, DLD, TD) shows a multivariate effect (F(4,178) = 20.238, $p < .001$, $\eta_p^2 = .313$). Follow-up ANOVAs for timed word and timed pseudoword reading were also significant, see right-hand columns of Table 1. Posthoc analyses indicate that the pattern of performance is Dyslexia < DLD < TD ($p < .01$).

With respect to language, we administered a receptive vocabulary task (Peabody Picture Vocabulary Test-III-NL, [PPVT-III-NL]; Schlichting, 2005), morphology and sentence repetition task (Clinical Evaluation of Language Fundamentals - Dutch version, [CELF-4-NL], Kort et al., 2008), see Instruments and Table 1. A MANOVA with group and the language outcomes shows a multivariate effect ($F(4, 178) = 14.529$, $p < .001$, $\eta_p^2 = .334$). Follow-up ANOVAs for the language measures separately are significant (right-hand columns of Table 1). Posthoc analyses establish that the pattern of performance for vocabulary and sentence repetition is DLD < Dyslexia, TD and for morphology (word structure) DLD < Dyslexia < TD.

Finally, the Raven Progressive Matrices subtest (Raven & Raven, 2003) was used to measure nonverbal intelligence. Children had to obtain a percentile score of at least 17%, a score in the lower bound of the normal range. The groups did not differ from each other on this measure, see right-hand columns of Table 1.

Table 1: Means and Standard Deviations of Reading and Language Outcomes per Group

Language measure	Dyslexia	DLD	TD	$F(2,89)$	η_p^2
Word reading (EMT)#	3.2 (2.1)	5.3 (3.7)	10.1 (2.4)	47.409**	.517
Pseudoword reading (Klepel) #	4.4 (1.5)	5.9 (3.1)	11.1 (2.3)	68.524**	.606
Receptive vocabulary (PPVT)%	56.4 $(19.7)_a$	32.3 (21.1)	58.2 $(23.6)_a$	13.714**	.236
Sentence repetition (CELF ZH)#	9.4 $(2.0)_a$	4.8 (2.3)	9.7 $(2.1)_a$	35.636**	.445
Word structure (CELF WS)@ !	26.5 (1.4)	21.9 (4.3)	27.8 (1.5)	38.736**	.465
Non-verbal IQ (Raven)$	57.6 (24.3)	60.3 (21.8)	58.7 (27.5)	.093	.002

Note. Subscripts in the group columns indicate that groups do not differ from each other significantly in case of posthoc testing. # standard score. % percentile score. $ norm score. @ raw score. ! results on the raw score remain the same when age is entered as a covariate. ** $p < .001$.

Instruments

Word spelling. A word dictation task was used to assess spelling (*Schoolvaardigheids-toets Spelling*, SVS, [School Skill Test Spelling]; de Vos & Braams, 2015). The task consists of 30 targets, divided in two blocks of 15 targets. Targets are embedded in sentences and repeated after each sentence. An example is: 'Dat poesje blijft erg klein. Schrijf op 'klein'.' ['That pussycat remains very small. Write down 'small'.']. The targets differ partly per grade: in grade 2, blocks 2 and 3 are presented; in grade 3 blocks 3 and 4 and in grade 4 blocks 4 and 5. Reliability of the (two versions of the) SVS is reported to be 0.82 or higher (Expertgroep Toetsen PO, 2016).

On the basis of the number of incorrectly spelled items (max 30), raw correct (max 30) and percentile scores were calculated. Furthermore, the number of incorrect items per error type was generated. The SVS software generates this output based on targets and realizations. The dictation tasks of the present sample contained 14 error categories, see Table 2 below for the categories and examples of errors.[7] We categorized these errors as being part of one of 5 main categories: 1) transparent phoneme-grapheme associations, 2) less

[7]The errors made by the groups in the categories of the SVS dictation task (see Method, Table 2) are displayed in Supplementary Figures 3 and 4.

transparent phoneme-grapheme associations, 3) phonology, orthography, morphology required, 4) spelling rules (requiring systematic operations on phonology, orthography, morphology), and 5) orthographic knowledge. These categories are presented in Table 2 (fourth column). In the transparent phoneme-grapheme category (1), an error is made when the mapping between the phoneme and grapheme is straightforward as there are no alternative spellings and the phoneme grapheme required reflects the phonology. In the less transparent phoneme-grapheme category (2), there are errors in targets in which a connection between phonemes and graphemes needs to be made that is complicated by homographs (e.g., ei/ij, au/ou) or by complex phonology (e.g., schr /sxr/). Category 3 refers to errors when a combination of phonology, orthography and morphology is required, such as spellings of bound morphemes. Category 4 refers to spelling rule errors, and thus to spelling rules that are taught explicitly, such as open and closed syllables (and vowel length) and final consonant spelling (e.g., Gills & Ravid, 2006). These also require integration of phonology, morphology and orthography. Finally, category 5 refers to errors related to orthographic knowledge and conventions (such as loanwords, hyphens, spacing).

Note that more than one error could occur in the 30 target items, for instance *pittig* [spicy] could be spelled as **piteg* and thus include both an error in open and closed syllables (*itt* as **it*, category 7) as well as in bound morphemes (*ig* as **eg*, category 11). The total number of errors was also tallied.

Literacy(-related) skills

Word reading. A timed word reading test (*Eén Minuut Test*, EMT, [One Minute Test]; Brus & Voeten, 1999) and a timed pseudoword reading test (*Klepel*; van den Bos et al., 1994) were used to evaluate word reading fluency. During the task, the child has one and two minutes, respectively, to read aloud as many (pseudo)words as possible. The number of syllables of the targets increases in length. The score is the number of correctly read (pseudo)words (maximum of 116 in each task). Standard scores were calculated on the basis of these raw scores.

RAN. RAN letters (a, d, o, p, s) was assessed with a subtest of the test for Continuous Naming and Word Reading (*Continu Benoemen & Woorden Lezen*; van den Bos & Lutje Spelberg, 2007). The children were shown a sheet with five columns of ten letters each and were asked to name all 50 letters as quickly and accurately as possible. The score on the RAN subtests is reported as the number of seconds required to complete the task: a higher score thus corresponds to weaker performance.

Language skills

Receptive vocabulary. The Dutch version of the Peabody Picture Vocabulary Test (PPVT-III-NL; Schlichting, 2005) was used to measure receptive vocabulary knowledge. Children were shown four pictures on the screen with an orally presented target word. They were asked to point to the picture that best resembled this target. Each set contained 11 items, with a total of 17 sets. Testing and scoring proceeded according to the manual. Raw scores were

Table 2: Spelling Error Categories (de Vos & Braams, 2015) (Column 2), Examples from Current Dataset (Column 3) and Categories Applied in the Current Study (Column 4)

Nr	Category (de Vos & Braams, 2015)	Examples of errors in present study	Categories in present study
1	Letters or digraphs that can be spelled only in one way	Error in digraph *oe* for /u/: *boek* as **beok* or **bok* [book]	Transparent phoneme-grapheme (1)
2	Letters and letter combinations with pronunciations that lead to incorrect spelling	Error in digraph nk for /ŋk/: *plank* as **plangk* [plank] Error in f/v: fatsoen as *vatsoen [decency] Error in /i:/ as i/ie: *ski* as **skie* [ski]	Less transparent phoneme-grapheme (2)
3	Vowel or consonant cluster	Error in eeuw/ieuw/uw: *schaduw* as *schaduuw* [shadow] Error in sch or schr: *ongeschikt* as **ongescrikt* [unsuited]	Less transparent phoneme-grapheme (2)
4	Non-transparent consonant cluster or digraph	Error in ou/au/ouw for /au/: *gauw* as **gouw* or **gou* [quick] Error in ei/ij for /ɛi/: *trein* as **trijn* [train]	Less transparent phoneme-grapheme (2)
5	Loanwords	Spelling *procent* as **prosent* [percent]	Orthographic knowledge (5)
6	Abbreviations, additions, symbols	Spelling *'s avonds* as* *s'avonds* [at night]	Orthographic knowledge (5)
7	Open and closed syllables (requiring either one grapheme for a long vowel or a double consonant)	Spelling *metro* as **meetro* [metro] Spelling *koffie* as **kofie* [coffee]	Spelling rule (Phonology-Orthography-Morphology) (4)
8	Compounds: as one word or separate, using a merging -s and hyphenation	Spelling *tekort* as **te kort* [deficit as too short] and *te kort* as **tekort* [too short as deficit]	Orthographic knowledge (5)
9	Vowel clash, demanding diaeresis, hyphen or being spelled as one	Spelling *skieën* as **skiejen* [to ski]	Orthographic knowledge (5)
10	Extension rules to discover underlying sound. In Dutch, final devoicing takes place, obscuring underlying voicing values (hond is pronounced /hOnt/, but the plural is honden /hOnd@n/	Spelling *neven* as **nefen* [cousins/nephews] Spelling *gemiddeld* as **gemiddelt* [average]	Spelling rule (Phonology-Orthography-Morphology) (4)
11	Bound morphemes (prefix, suffix) and silent vowels	*pittig* as **pitteg* [spicy]	Phonology-Orthography-Morphology (3)
12	Diminutives. In Dutch, these are dependent on the target stem (-je, tje, kje, pje, nkje)	Spelling *kettinkje* as **kettingje* [small necklace]; *fotootje* as **fototje* [small photo]	Spelling rule (Phonology-Orthography-Morphology) (4)
13	Suffix of loanwords (-eit, -isch)	Spelling *majesteit* as **majestijd* ([ajesty]	Orthographic knowledge (5)
14	Contractions spelled with a middle –e-	Spelling *onafscheidelijk* as **onafschijdlijk* [inseparable]	Phonology-Orthography-Morphology (3)

converted to percentile scores.

Sentence repetition was assessed through the Dutch version of the CELF-4 recalling sentences task (CELF-4-NL; Kort et al., 2008). In this task the experimenter presents a sentence and the child is asked to repeat it. The sentences increase in length and complexity as the task progresses. Testing and scoring proceeded according to the manual. Raw scores were converted to standard scores.

Morphology was evaluated through the Dutch version of the CELF-4 word structure task (CELF-4-NL; Kort et al., 2008). In this task, the experimenter presents a sentence and a picture to the child. The child is required to complete the sentence on the basis of the target frame of the experimenter and the picture, such as 'Deze jongen staat en deze jongen…[zit]' ['This boy stands and this boy… [sits]']. Different types of morphology are assessed, including pronouns, nouns (diminutives and plurals), verbs (tense, compound verbs, subject-verb agreement), and adjectives (comparatives and superlatives). Responses were coded as (in)correct and the total raw score was calculated (max 30). These scores were not transformed to percentile scores, as these norm scores are available only up to 8 years.

Background variables: General cognitive skills

Non-verbal IQ was assessed through Raven's Standard Progressive Matrices (Raven & Raven, 2003), which is a standardized measure of nonlinguistic intelligence. Raw scores were converted to percentile scores.

Memory skills. Short-term and working memory were assessed in the verbal and in the visuospatial domain. For the verbal domain, we used the standardized forward and backward digit span tasks in the CELF-4-NL (Kort et al., 2008). Testing and scoring proceeded according to the manual. In the forward version of the task, children had to repeat sequences of digits that increased between 2 and 9 digits in the correct order. In the backward version, children had to repeat the sequences in the reversed order. Raw scores were converted to standard scores. For the visuospatial domain, we used a non-standardized Dutch adaptation of the Dot Matrix task in the Alloway Working Memory Assessment (AWMA; Alloway, 2007) that also contained a forward and backward version. Children were shown a matrix (4 x 4) in which dots appeared according to sequences of increasing length. After presentation, children had to point to the locations on the matrix where the dots had appeared (forward version: in the same order, backwards version: in the reversed order). Testing and scoring proceeded according to the manual. The maximum score was 36 (six trials per block and a maximum of six blocks).

Sustained attention was measured using the Dutch version of the Score! Subtest of the Test of Everyday Attention for Children (TEA-Ch; Schittekatte et al., 2007). This is an auditory task in which children listen to 10 audio fragments that contain between 9 and 15 target sounds. The child's task is to (silently) count the number of target sounds per fragment, taken to reflect their ability to maintain attention over time. Raw scores represent the number of items answered correctly out of 10.

Procedure

This sample of children was part of a larger project, Progracy, in which comparisons between children with dyslexia and TD children and children with DLD and TD children were made. This larger project focused on the relation between statistical learning and language proficiency: children that participated also performed three statistical learning tasks in addition to the tasks reported on in this paper. The full test battery took two to four sessions of one hour each, spread over 2 to 3 weeks for each child. Each test session started with a statistical learning task and was then followed by a set of cognitive and language measures. The sessions consisted of the following clusters: 1) word reading, verbal and visuospatial memory tasks and a nonadjacent dependency statistical learning task, 2) sentence recall, word structure, sustained attention, non-verbal intelligence tasks, and a serial reaction time task, and

3) spelling, vocabulary, RAN tasks and a visual statistical learning task. The order in which participants performed the different sessions was counterbalanced. The study has been approved by the ethics committee of the Faculty of Humanities of the University of Amsterdam in 2016.

Results

Q1: Mean spelling outcomes

Prior to group comparisons on the spelling outcomes, data screening was conducted. There were no outliers (±3.27 SD from the mean on the spelling score). The data were not normally distributed for the percentile score (left-skew; $W(50) = .929$, $p = .005$) but were for the raw score ($W(50) = .971$, $p = .248$) and the total number of errors ($W(50) = .973$, $p = .308$).

The mean percentile scores confirm the poor performance of the dyslexia group ($M = 9.1$, $SD = 12.7$) and the average performance of the TD group ($M = 46.7$, $SD = 24.9$). They indicate that mean performance of the DLD is poor ($M = 11.10$, $SD = 14.6$). The raw spelling scores and the total number of errors are presented in Table 3, along with statistical testing outcomes (ANOVAs). For the raw scores, performance is ranked: TD > Dyslexia, DLD. For the total number of errors it is TD > Dyslexia > DLD.

On the basis of the word reading outcomes, we found that 18/30 children with DLD (60%) displayed poor reading (standard score ≤6) on either or both word reading tasks. They were classified as DLD+reading difficulties (RD). The mean spelling percentile score of the DLD+RD group was very low (M = 4.00, SD = 4.8). For the DLD-only group it was 21.8 (SD = 17.9). The raw spelling outcomes were subsequently analyzed through ANOVAs for the four groups (DLD+RD, DLD-only, Dyslexia, and TD), see lower section Table 3. The pattern of outcomes for the raw score and total error score is TD > DLD-only, Dyslexia > DLD+RD. In sum, the spelling scores indicate that spelling outcomes are poor and delayed for both the DLD and dyslexia groups, and, within the DLD-group, most clearly for the DLD+RD group.

Table 3: Mean Spelling Outcomes (SD) per Group and Statistical Outcome

Outcome	Dyslexia		DLD	TD	$F(2,89)$	η_p^2
Raw score (max 30)	6.87 (3.7)a		6.27 $(4.9)_a$	18.6 (4.9)	72.790**	.621
Number of errors#	36.29 (2.4)		44.50 (2.4)	17.06 (2.4)	34.481**	.437
	Dyslexia	DLD+RD	DLD-only	TD		
Raw score (max 30)	6.87 $(3.7)_a$	3.61 (3.0)	10.25 (4.3)a	18.6 (4.9)	64.258**	.687
Number of errors#	36.29 $(2.4)_a$	53.06 (14.2)	31.67 $(9.8)_a$	17.06 (2.4)	36.573**	.555

Note. Subscripts indicate that groups do not differ from each other significantly. ** $p < .001$. # the number of errors of all categories added up.

Q2: Spelling errors

Figures 1 and 2 present the number of errors for each of the five spelling categories for the three groups (Figure 1: Dyslexia, DLD, TD) and four groups (Figure 2: Dyslexia, DLD-only, DLD+RD, TD).7 Generally, the number of errors is highest for errors related to spelling rules (category 4) and phonology-orthography-morphology-mappings (category 3). For the three groups (TD, DLD, Dyslexia), the pattern of performance for all different categories is TD > Dyslexia > DLD. For the four groups, this is TD > DLD > Dyslexia, DLD+RD. Note that the DLD+RD group makes more errors in all error categories, including errors in transparent phoneme-grapheme mappings (category 1) and phonology-orthography mappings (category 2).

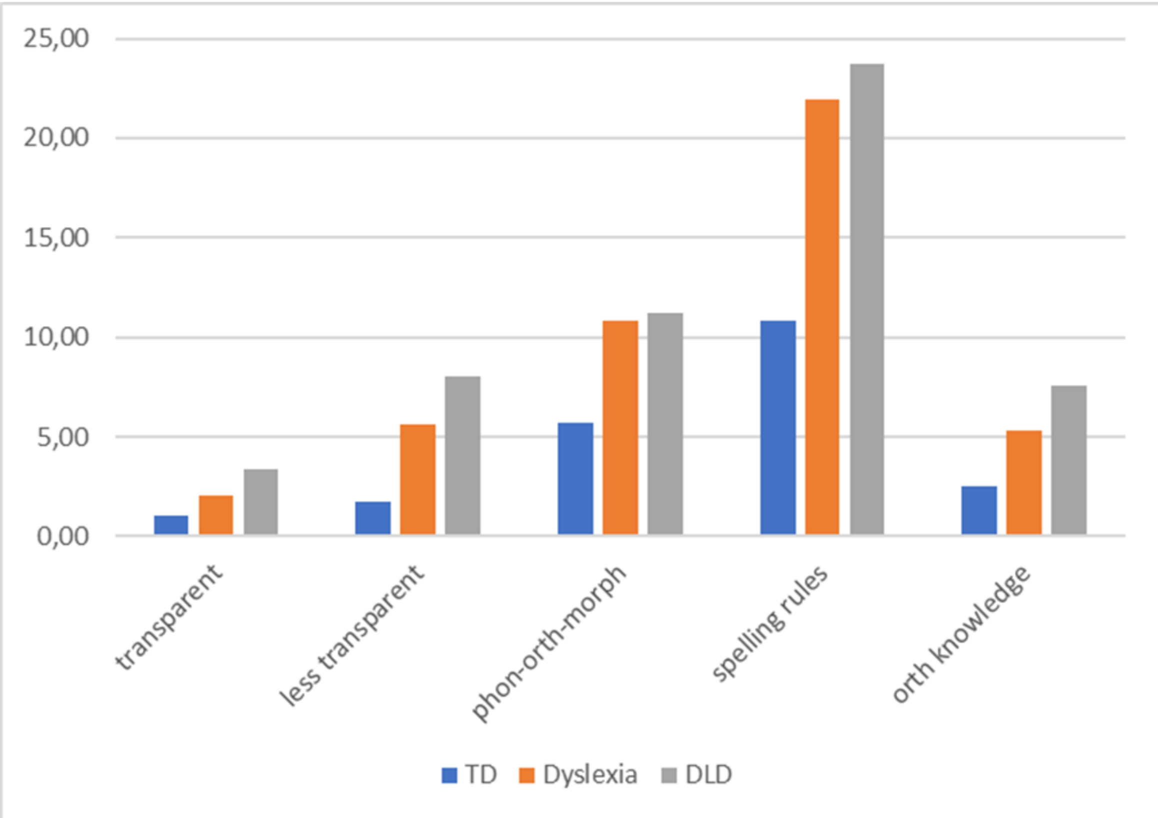


Figure 1: Number of Errors by Spelling Category (Table 2. Column 4) for Three Groups

A MANOVA with the five different error categories and group (Dyslexia, DLD, TD) shows a multivariate effect (Wilk's Lambda (10, 170) = 7.427, $p < .001$, $\eta_p^2 = .304$). Follow-up ANOVAs show significant effects on all five categories: (1) transparent phoneme-graphemes $F(2, 89) = 13.353$, $p < .001$, $\eta_p^2 = .231$, (2) less transparent phoneme-graphemes $F(2, 89) = 28.296$, $p < .001$, $\eta_p^2 = .389$, (3) phonology-orthography-morphology $F(2, 89) = 10.185$, $p < .001$, $\eta_p^2 = .186$, (4) spelling rules $F(2, 89) = 19.770$, $p < .001$, $\eta_p^2 = .308$, and (5) orthographic knowledge

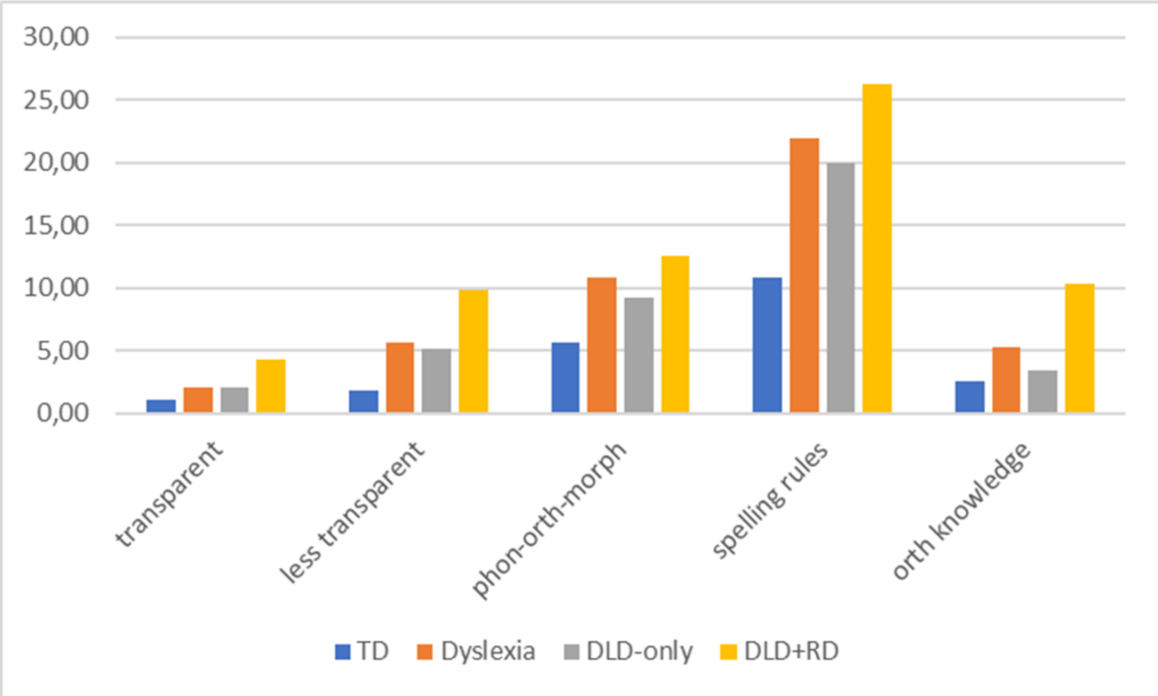


Figure 2: Number of Errors by Spelling Category (Table 2. Column 4) for Four Groups

$F(2, 89) = 7.525$, $p = .001$, $\eta_p^2 = .145$). Posthoc comparisons (all Games-Howell) show that for the errors in transparent mappings (category 1), the pattern is TD > Dyslexia > DLD; for the other four categories, this is TD > Dyslexia, DLD. Thus, the control group always makes fewer errors and the DLD and dyslexia groups resemble each other in the number of spelling errors that require more than transparent phoneme-grapheme mapping.

When the MANOVA is conducted for the four groups (Dyslexia, DLD-only, DLD+RD, TD), there is also a multivariate effect (Wilk's Lambda (15, 258) = 7.262, $p < .001$, $\eta_p^2 = .298$) and there are main effects for all categories separately: (1) transparent phoneme-graphemes $F(3, 92) = 14.062$, $p < .001$, $\eta_p^2 = .324$, (2) less transparent phoneme-graphemes $F(3, 92) = 28.277$, $p < .001$, $\eta_p^2 = .491$, (3) phonology-orthography-morphology $F(3, 92) = 7.820$, $p < .001$, $\eta_p^2 = .210$, (4) spelling rules $F(3, 92) = 14.943$, $p < .001$, $\eta_p^2 = .337$, and (5) orthographic knowledge $F(3, 92) = 10.486$, $p < .001$, $\eta_p^2 = .270$). On the categories less transparent phonology-orthography (2), phonology-orthography-morphology (3) and spelling rules (4), posthoc comparisons (all Games-Howell) show that there are no indications of differences between the three clinical groups (Dyslexia, DLD-only, DLD+RD). The TD group outperforms the three groups on these three categories. In terms of errors in transparent phoneme-grapheme mappings (category 1), the TD group makes fewer errors than the dyslexia and DLD+RD groups. Furthermore, the DLD+RD group makes more errors than the Dyslexia group. There are no indications that the DLD-only group differs from the dyslexia group and from the DLD+RD group. For the orthographic knowledge errors (5), the TD group makes fewer errors than the Dyslexia and DLD+RD group and the DLD+RD group makes significantly more errors than the DLD-only group.

In sum, the three group comparison (Dyslexia, DLD, TD) shows that the TD group makes fewer errors of each category. The two clinical groups resemble each other in the types of errors made, except that the DLD group makes more errors with transparent phoneme-grapheme correspondences (category 1). For the four group comparison, there was no evidence for or against a difference between the three clinical groups (Dyslexia, DLD-only and DLD+RD), as they show a similar number of errors of less transparent phoneme-grapheme mappings (category 2) and phonology-orthography-morphology associations (category 3) and spelling rules (category 4). For errors in transparent phoneme-grapheme mappings (category 1) and orthography (category 5), the DLD+RD-group shows most difficulties.

Q3: Factors contributing to spelling performance

In order to evaluate whether literacy, language, and general cognitive skills contribute to the spelling outcomes, regression analyses were conducted with raw spelling score as dependent variable. Group was not entered as a separate variable, as there is no a-priori assumption that variables contributing to spelling performance differ between the groups (Werfel et al., 2021).

The independent variables focused on literacy (timed word reading and RAN), language (receptive vocabulary, sentence repetition, and morphology), and general cognitive ability (sustained attention, verbal memory, visuospatial memory, and non-verbal IQ). Mean outcomes on these variables are presented in Table 4. With respect to timed word reading, only the real-word reading task was included, as there was a strong correlation between the real and pseudoword reading task, both on raw $r(92) = 0.92$, $p < .001$ and standard scores $r(92) = 0.89$, $p < .001$.

Data screening for the 11 independent variables yielded one significant multivariate outlier (using Mahalanobis distance calculations), with the probability of multivariate outliers set at $t < .01$. As this multivariate outlier came from a clinical group (DLD), the outlier status was not considered problematic and the data were included in the sample.

Correlations between the independent variables are presented in Table 4 and correlations with the raw spelling score in Table 5. There are strong and significant correlations between timed word reading and RAN, and between the language measures (vocabulary, morphology and sentence repetition). As they are still all below the .7 value, all variables were retained in the regression analyses. Furthermore, as sustained attention, visuospatial memory and non-verbal IQ did not correlate with spelling outcomes, these background variables were excluded from the regression analyses.

A subsequent regression analysis with the spelling correct score was significant ($F(7, 91) = 28.522$, $p < .001$) and explained 68% of the variance (adjusted $R^2 = .679$), see Table 6. Timed word reading and RAN were the only unique significant contributors; the language measures were not. This remained the case when only one of the language measures was included, meaning that this finding cannot be due to collinearity of language measures. An exploratory analysis was conducted to investigate the relationship between oral language

Table 4: Mean Outcomes and Correlations between the Independent Variables

	Mean (SD)	RAN	Vocab	SRep	Morph	SA	VSTM	VWM	VisSTM	VisWM	NV-IQ
Word reading	43.1 (19.7)	-.645**	.278**	.365**	.461**	.219*	.447**	.232*	.235*	.276**	.146
RAN letters	33.8 (10.3)	1	-.177*	.-180*	-.209*	.249**	-.316**	-.175*	-.359**	-.229*	-.056
Vocabulary	110.8 (11.2)		1	.691**	.676**	.199*	.299**	.308**	.352**	.221*	.210*
Sentence Rep	48.6 (16.8)			1	.734**	.126	.539**	.451**	.311**	.258*	.274*
Morphology	25.5 (3.7)				1	.222*	.411**	.350**	.224*	.191*	.162
Sustained att	7.18 (2.3)					1	.178*	.115	.166	.337**	.154
VSTM	7.5 (1.8)						1	.371**	.162	.215*	.145
VWM	3.9 (1.3)							1	.254**	.211*	.312**
Visspat STM	23.9 (4.8)								1	.539**	.338**
Visspat WM	22.2 (5.1)									1	.420**
NV-IQ	36.2 (6.5)										1

Note. ** $p < .01$ * $p < .05$

Table 5: Correlations between the Raw Spelling Score and the Independent Variables (Raw Scores)

	Spelling
Word reading	.803**
RAN	-.413**
Vocabulary	.294**
Morphology	.414**
Sentence repetition	.483**
Sustained attention	.050
Verbal STM	.503**
Verbal WM	.289**
Visuospatial STM	.152
Visuospatial WM	.183
Non-verbal IQ	.087

Note. ** $p < .001$.

ability and spelling. When the regression analysis was conducted with the variables excepting word reading, morphology ($B = .369$, $t = 2.800$, $p = .006$) as well as VSTM ($B = .297$, $t = 2.825$, $p = .006$) did provide unique contributions to the spelling outcomes in a significant regression model ($F(6, 91) = 9.670$, $p < .001$), although the explained variance was lower than in the model that did contain word reading (adjusted $R^2 = .364$).

The pattern of findings was similar when the total number of errors was the dependent variable.

Table 6: Standardized Beta Coefficients of Independent Variables Predicting Spelling Outcomes

All raw scores	Spelling raw score (max 30) Standardized coefficients Beta	*t*
Timed word reading	.811	9.197***
RAN letters	.177	2.247**
Vocabulary	-.013	-.146
Sentence repetition	-.003	-.025
Morphology	.076	.770
VSTM	.151	1.983
VWM	.052	.776

Note. *** $p < .001$. ** $p < .01$.

Discussion

The goal of this study was to gain insight into the similarities and differences of the spelling of children with dyslexia and DLD and the role that literacy and language abilities play in their spelling outcomes. We targeted three questions to address this issue, namely 1) whether children with DLD, dyslexia, and typically developing children show differences in spelling dictation outcomes, 2) whether the pattern of errors differs between the dyslexia and DLD groups (and the subgroups of DLD-only and DLD+reading difficulties (RD), and 3) whether spelling outcomes are predicted by both literacy(-related) abilities as well as language abilities.

Spelling outcomes

The three mean spelling outcomes (percentile scores, raw correct score and total error score) indicated that the TD group outperformed the Dyslexia and DLD groups on spelling and that the latter two groups show low spelling outcomes. Within the DLD-group, this was most clearly the case for the children with DLD+ RD.

The spelling difficulties for the group of children with dyslexia were anticipated, as reading, a core deficit in any definition of dyslexia, is related to spelling (Georgiou et al., 2020), and as poor spelling can be part of the dyslexia diagnosis (in the Netherlands). Children with DLD also showed poorer spelling outcomes than their age-matched peers, in line with previous findings (meta-analyses by Joye et al., 2019; Graham et al., 2020).

The finding that for the DLD group the spelling difficulties were most clearly present for the subgroup of children with DLD+RD is similar to previous findings (Bishop et al., 2009; McCarthy et al., 2012; Snowling et al., 2019; Vandewalle et al., 2012). In our study, the spelling percentile score of the DLD+RD group was extremely low (4th percentile). Also, the mean raw spelling score and total number of errors were even lower than that of the dyslexia-only group. These children thus have severe spelling difficulties. The division between the children with DLD-only and children with DLD+RD speaks to the literature, which has shown that DLD and severe word reading difficulties can be co-occurring and comorbid disorders, but that DLD as such does not automatically entail a literacy disorder (e.g., Adlof & Hogan, 2018; Catts et al., 2005; Rakhlin et al., 2013; Snowling et al., 2019).

Error types

With respect to the types of errors, we found that the control group made fewer errors in all error categories than the dyslexia and DLD groups. These categories were 1) transparent phoneme-grapheme associations, 2) less transparent phoneme-grapheme associations, 3) phonology, orthography, morphology, 4) spelling rules, and 5) orthographic knowledge. We found no evidence that the children with dyslexia and children with DLD groups differed from each other, except for the errors in transparent phoneme grapheme mappings, where the group of children with DLD made more errors than the group of children with dyslexia.

The error comparison between the four groups did not provide evidence for or against a difference between the three clinical groups (Dyslexia, DLD-only and DLD+RD) as they made a similar number of errors on three of the five categories (less transparent phoneme-grapheme mappings and phonology-orthography-morphology associations, spelling rules). The errors made in these categories furthermore suggest that all three clinical groups make errors when spelling is not straightforward and requires more complex associations or rules. Furthermore, the DLD+RD and Dyslexia groups made more errors than the control group on straightforward phoneme-grapheme mappings, indicating that for these groups, the basis phoneme-grapheme associations are challenging. Finally, these groups also made more errors than the control group regarding orthographic knowledge.

The findings thus indicate that children with dyslexia make errors when connections between phonology-orthography and phonology-orthography-morphology (including spelling rules) need to be made. This is in line with the literature (Angelelli et al., 2004; Landerl & Wimmer, 2000; Protopapas et al., 2013), and points to the need of systematic spelling instruction and treatment for these children, even at the most basic phoneme-grapheme associations. The errors made by the children with DLD align with studies that point to phonological errors (e.g., Broc et al., 2013; Joye et al., 2020; Larkin & Snowling, 2008; McCarthy et al., 2012), as well as those that have reported orthographic and morphological errors in this group (Deacon et al., 2014; Joye et al., 2020; McCarthy et al., 2012; Silliman et al., 2006; Windsor et al., 2000; Wolter & Apel, 2010), especially in more transparent orthographies (see also Broc et al., 2021). They indicate that children with DLD also have a need for systematic and intensive spelling instruction. As the DLD group showed most difficulties with straightforward phoneme-grapheme mappings, it is clear that they require intensive spelling instruction from the outset of literacy instruction. Across-the-board spelling difficulties were present mainly for the DLD+RD group: this group requires the greatest amount of support, as spelling performance was extremely low and spelling errors were evident in all error categories. Basic letter-sound associations need to be strengthened for these children, as do orthographic representations of specific words, as well as integrations of orthography and linguistic rules.

In order to understand the mechanisms of spelling errors more clearly, it could be investigated which specific sources of information children use during spelling specifically selected target types and orthographic rules/patterns. This could be done by assessing dictation of words that require specific morphological, lexical, phonological and orthographic sources of knowledge (e.g., de Bree et al., 2017; Kim et al., 2016). In this approach, words are selected that require specific phonological, morphological and orthographic operations (i.e., spelling plurals such as taken for singular taak [task] and takken for singular tak [branch], or past tense leefde for infinitive leven [live]). These words also differ in their lexical frequency and age of acquisition. Such a comparison can indicate whether specific measures of language, orthography and frequency contribute similarly to the spelling outcomes for the different groups. Gaining insight in the types of errors across specific targets can determine most clearly which instruction is most suitable.

Language and literacy as predictors of spelling

With respect to the third research question, whether literacy and language abilities predicted spelling outcomes, it was found that word reading was the dominant contributor to spelling outcomes. RAN was an additional significant unique predictor, whereas we found no evidence that language abilities and general cognitive abilities predict (or do not predict) spelling outcomes. However, if word reading was left out of the analyses, morphology and VSTM became unique significant contributors. Note that this coincided with a decreased amount of explained variance. If only the language measure morphology was included (and leaving out vocabulary and sentence repetition), then morphology did contribute to the model of total errors. There are thus subtle indications of contributions of language measures to spelling. Nevertheless, the regression analyses generally pointed to the importance of reading measures for spelling (Georgiou et al., 2020; Werfel et al., 2021).

These findings match other studies that report a contribution of language to spelling (Joye et al., 2019) but especially or only if reading is not included in the analyses (Joye et al., 2019; Kim et al., 2013; Larkin et al., 2013; McCarthy et al., 2012). Word reading might subsume effects of language variables. This was also visible in the correlations between word reading and the language measures in our sample. Reading ability requires linguistic knowledge, as decoding and sight word reading is facilitated by vocabulary and morphology. The findings also speak to those of van Weerdenburg et al. (2011), who found that verbal-sequential processing had the highest predictive value of spelling of all language measures included. This factor contained RAN, which was found to be a strong contributor to spelling in our sample.

With respect to RAN, we had not necessarily anticipated a contribution of this measure to spelling above and beyond word reading. RAN can be taken to be a microcosmos of reading (Norton & Wolf, 2012), as it requires conversion of visual information to an auditory response as fluently as possible. Furthermore, RAN is a measure determined by speed, whereas spelling is not a timed measure. This continued contribution of RAN to spelling might indicate that RAN taps the efficiency of the alphabetic principle (Moll et al., 2009), which is thus poorer in children with dyslexia and children with DLD than their age-matched peers.

Limitations

This study is qualified by some important limitations. First, the sample size is limited, most prominently for the subgroups DLD-only (n = 12) and DLD+RD (n = 18). The findings can therefore not be taken to generalize to the populations of DLD, dyslexia and typically developing children at large, even though the pattern of findings largely resembles previous findings related to spelling performance in DLD and dyslexia, spelling errors in dyslexia, and contributors to spelling performance.

Related, the comparison now only consists of the spelling of monolingual children or children with a Dutch-dominant background. As bilingualism is becoming increasingly frequent, future research should also include bilingual speakers in the assessment of spelling and contributions of language and word reading (e.g. de Bree et al., 2022).

Furthermore, the contribution of language and literacy to spelling outcomes is determined on the basis of concurrent rather than longitudinal data. Longitudinal studies are necessary to evaluate whether earlier language abilities contribute more clearly to later spelling outcomes in children with (a risk of) literacy and language abilities than those measured concurrently, or whether mediating effects can be attested (Georgiou et al., 2020; Lervåg & Hulme, 2010; van Weerdenburg et al., 2011; Werfel et al., 2021).

Fourth, the dictation presented to the children was based on the grade they were attending. As the children did not all attend the same grade, the targets differed somewhat. This means that the errors that were made might also have differed due to some of the targets differing between the children. Thus, although all children attended grades in which spelling rules and associations between phonology-orthography and morphology are part of the spelling curriculum, a more controlled assessment of spelling is needed to further inform our understanding of the spelling acquisition of the different groups of children. A study with this approach is being conducted for spelling of children with DLD (Bliekendaal et al., in preparation). It can inform us more about the areas of spelling difficulty these children have, and also which instruction and which interventions might be helpful.

Fifth, we have not taken into account the spelling instruction the children have received. Spelling instruction is essential for spelling development (Graham & Santangelo, 2014), but schools can differ in the quality of spelling instruction as well as in the mean spelling outcomes obtained by their students (e.g., Cordewener et al., 2012). Furthermore, we know that specific spelling interventions for poor spellers/readers lead to improvement (Galuschka et al., 2014, 2020). Therefore, the amount and type of instruction could have impacted on the outcomes. Children with diagnosed dyslexia will have received or have been receiving specialized and intensive literacy interventions both at special dyslexia centres as well as at school. The fact that they still show poor performance on spelling points to the severity of their literacy deficit and that continued support is required. It is also expected that children with DLD in general will have had more intensified spelling instruction at school, as it is common practice to provide these children with additional literacy instruction, given their increased risk of difficulties in this area. However, it is not clear whether the amount and type of instruction was the same for all children. For the subgroup of children with both DLD and (a suspicion of) dyslexia, more intensified instruction at school and specialist dyslexia treatment will have commenced. However, little is known about the treatment outcomes of children with comorbid disorders, such as dyslexia and DLD. It is evident that the absence of the information on the amount and type of spelling instruction and treatment limits the interpretations of the current study and requires further study. Future studies should thus connect spelling outcomes both to item characteristics (which targets assessed, frequencies, age of acquisition), child characteristics (language, word reading), as well as to school and treatment characteristics (e.g., methods used, quality of instruction, additional intervention received). Such a project is currently being conducted in regular school settings (Drijver et al., in preparation).

A final limitation is that we only compared spelling of the clinical groups to age-matched peers, not younger language-or literacy-matched control groups. This was not the aim of our study but might have provided insight in the amount of delay as well as potential similarities between the younger peers in spelling errors of the clinical groups.

Conclusions

Our findings indicate that children with dyslexia and children with DLD show spelling difficulties, and children with DLD and additional reading difficulties do so most prominently. The spelling errors show that errors arise when phonology-orthography and morphology need to be associated and rules have to be applied. For children with dyslexia and DLD+RD, errors are also more prominent in more basic phoneme-grapheme associations. These findings warrant (continued) intensive and systematic spelling intervention. The finding that there is a strong influence of reading ability on spelling outcomes and a less pronounced direct role of oral language underscores the importance of literacy instruction for spelling development and calls for the support of literacy skills as well as oral language in children at-risk of spelling difficulties.

References

Adlof, S. M., Baron, L. S., Bell, B. A., & Scoggins, J. (2021). Spoken word learning in children with developmental language disorder or dyslexia. *Journal of Speech, Language, and Hearing Research, 64*(7), 2734-2749. https://doi.org/10.1044/2021_JSLHR-20-00217

Adlof, S. M., & Hogan, T. P. (2018). Understanding dyslexia in the context of developmental language disorders. *Language, Speech, and Hearing Services in Schools, 49*(4), 762–773. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0049

Alloway, T. P., Tewolde, F., Skipper, D., & Hijar, D. (2017). Can you spell dyslexia without SLI? Comparing the cognitive profiles of dyslexia and specific language impairment and their roles in learning. *Research in Developmental Disabilities, 65*, 97-102. https://doi.org/10.1016/j.ridd.2017.04.013

Alloway, T. P. (2007). *Alloway Working Memory Assessment* (AWMA). Pearson. American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Angelelli, P., Judica, A., Spinelli, D., Zoccolotti, P., & Luzzatti, C. (2004). Characteristics of writing disorders in Italian dyslexic children. *Cognitive & Behavioral Neurology, 17*, 18–31

Apel, K., Wilson-Fowler, E. B., Brimo, D., & Perrin, N. A. (2012). Metalinguistic contributions to reading and spelling in second and third grade students. *Reading and Writing, 25*, 1283–1305. https://doi.org/10.1007/s11145-011-9317-8

Berninger, V. W. & Swanson, H. L. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. *Advances in Cognition and Educational Practice, 2*, 57–81.

Bishop, D. V. M., McDonald, D., Bird, S., & Hayiou-Thomas, M. E. (2009). Children who read words accurately despite language impairment: Who are they and how do they do it? *Child Development, 80*(2), 593–605. https://doi.org/10.1111/j.1467-8624.2009.01281.x

Bishop, D. V. M. & Clarkson, B. (2003). Written language as a window into residual language deficits: A study of children with persistent and residual speech and language impairments. *Cortex, 39*(2), 215–237. https://doi.org/10.1016/S0010-9452(08)70106-0

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & 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*(10), 1068-1080. https://doi.org/10.1111/jcpp.12721

Brizzolara, D., Gasperini, F., Pfanner, L., Cristofani, P., Casalini, C., & Chilosi, A. M. (2011). Long-term reading and spelling outcome in Italian adolescents with a history of specific language impairment. *Cortex, 47*(8), 955–973. https://doi.org/10.1016/j.cortex.2011.02.009

Broc, L., Joye, N., Dockrell, J.E. & Olive, T. (2021). Capturing the nature of the spelling errors in Developmental Language Disorder: A scoping review. *Language, Speech and Hearing Services in Schools, 52*, 1127–1140. https://doi.org/10.1044/2021_LSHSS-20-00086

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research, 80*(2), 144–179. https://doi.org/10.3102/0034654309359353

Broc, L., Bernicot, J., Olive, T., Favart, M., Reilly, J., Quémart, P., & Uzé, J. (2013). Lexical spelling in children and adolescents with specific language impairment: Variations with the writing situation. *Research in Developmental Disabilities, 34*(10), 3253–3266. https://doi.org/10.1016/j.ridd.2013.06.025

Brus, B. T., & Voeten, M. J. M. (1999). Eén minuut test [One minute test]. Hartcourt Test Publishers.

Caravolas, M. & Volín, J. (2001). Phonological spelling errors among dyslexic children learning a transparent orthography: The case of Czech. *Dyslexia, 7*(4), 229–245. https://doi.org/10.1002/dys.206

Casalis, S., Colé, P., & Sopo, D. (2004). Morphological awareness in developmental dyslexia. *Annals of Dyslexia, 54*(1), 114–138. https://doi.org/10.1007/s11881-004-0006-z

Catts, H. W., Adlof, S. M., Hogan, T. P., & Ellis Weismer, S. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research, 48*(6), 1378–1396. https://doi.org/10.1044/1092-4388(2005/096)

Cordewener, K. A. H., Bosman, A. M. T., & Verhoeven, L. (2012). Specific language impairment affects the early spelling process quantitatively but not qualitatively. *Research in Developmental Disabilities, 33*(4), 1041–1047. https://doi.org/10.1016/j.ridd.2012.01.011

Deacon, S. H., Cleave, P. L., Baylis, J., Fraser, J., Ingram, E., & Perlmutter, S. (2014). The representation of roots in the spelling of children with specific language impairment. *Journal of Learning Disabilities, 47*(1), 13–21. https://doi.org/10.1177/0022219413509965

de Bree, E.H., Boerma, T., Hakvoort, B., Blom, E., & van den Boer, M. (2022). Word reading in monolingual and bilingual children with developmental language disorder. *Learning and Individual Differences, 98*. https://doi.org/10.1016/j.lindif.2022.102185.

de Bree, E., van der Ven, S., & van der Maas, H. (2017). The voice of Holland: Allograph production in written Dutch past tense inflection. *Language Learning and Development, 13*(3), 215–240. : http://dx.doi.org/10.1080/15475441.2016.1217777

Devonshire, V. & Fluck, M. (2010). Spelling development: Fine-tuning strategy-use and capitalising on the connections between words. *Learning and Instruction, 20*(5), 361-371. https://doi.org/10.1016/j.learninstruc.2009.02.025

de Vos, T. & Braams, T. (2015). *Schoolvaardigheidstoets Spelling (School Skill Test Spelling).* Boom test uitgevers.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56*(8), 848-856. https://doi.org/10.1111/jcpp.12378

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders, 20*(3), 19–36. https://doi.org/10.1097/00011363-200020030-00005

Eisenmajer, N., Ross, N., & Pratt, C. (2005). Specificity and characteristics of learning disabilities. *Journal of Child Psychology and Psychiatry, 46*(10), 1108–1115. https://doi.org/10.1111/j.1469-7610.2004.00394.x

Elbro, C. & Arnbak, E. (1996). The role of morpheme recognition and morphological awareness in dyslexia. *Annals of Dyslexia, 46*(1), 209–240. https://doi.org/10.1007/BF02648177

Expertgroep Toetsen PO (2016). https://www.expertgroeptoetsenpo.nl/

Fowler, A. E. & Liberman, I. Y. (1995). The role of phonology and orthography in morphological awareness. In L. B. Feldman (Ed.), *Morphological aspects of language processing.* Lawrence Erlbaum.

Galuschka, K., Görgen, R., Kalmar, J., Haberstroh, S., Schmalz, X., & Schulte-Körne, G. (2020). Effectiveness of spelling interventions for learners with dyslexia: A meta-analysis and systematic review. *Educational Psychologist, 55*(1), 1-20. https://doi.org/10.1080/00461520.2019.1659794

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE, 9*(2), Article e89900. https://doi.org/10.1371/journal.pone.0089900

Georgiou, G. K., Torppa, M., Landerl, K., Desrochers, A., Manolitsis, G., de Jong, P. F., & Parrila, R. (2020). Reading and spelling development across languages varying in orthographic consistency: Do their paths cross? *Child Development, 91*(2), e266-e279. https://doi.org/10.1111/cdev.13218

Gillis S, & Ravid D. (2006) Typological effects on spelling development: a crosslinguistic study of Hebrew and Dutch. *Journal of Child Language, 33*(3): 621-59. doi:10.1017/s0305000906007434.

Goodwin, A. P. & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading, 17*(4), 257-285. https://doi.org/10.1080/10888438.2012.689791

Graham, S., Hebert, M., Fishman, E., Ray, A. B., Rouse, A. G. (2020). Do children classified with specific language impairment have a learning disability in writing? A meta-analysis. *Journal of Learning Disabilities, 53*(4), 292-310. https://doi.org/10.1177/0022219420917338

Graham, S. & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing, 27*(9), 1703–1743. https://doi.org/10.1007/s11145-014-9517-0

Graham, S., Harris, K. R., & Hebert, M. (2011). It is more than just the message: Analysis of presentation effects in scoring writing. *Focus on Exceptional Children, 44*, 1–12. doi: 10.1201/b10962-2

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89*(1), 170-182. https://doi.org/10.1037/0022-0663.89.1.170

Hoefflin, G. & Franck, J. (2005). Development of spelling skills in children with and without learning disabilities. *L1-Educational Studies in Language and Literature, 5*(2), 175–192. https://doi.org/10.1007/s10674-005-0917-6

Joye, N., Broc, L., Olive, T., & Dockrell, J. (2019). Spelling performance in children with developmental language disorder: A meta-analysis across European languages. *Scientific Studies of Reading, 23*(2), 129-160. https://doi.org/10.1080/10888438.2018.1491584

Joye, N., Dockrell, J. E., & Marshall, C. R. (2020). The spelling errors of French and English children with developmental language disorder at the end of primary school. *Frontiers in Psychology, 11*(1789), 1-21. https://doi.org/10.3389/fpsyg.2020.01789

Kort, W., Compaan, E., Schittekatte, M., & Dekker, P. (2008). *Clinical evaluation of language fundamentals (CELF-4-NL). Nederlandse versie. Handleiding.* Pearson Assessment and Information.

Kim, Y.-S. G., Petscher, Y., & Park, Y. (2016). Examining word factors and child factors for acquisition of conditional sound-spelling consistencies: A longitudinal study. *Scientific Studies of Reading, 20*(4), 265–282. https://doi.org/10.1080/10888438.2016.1162794

Kim, Y.-S., Apel, K., & Al Otaiba, S. (2013). The relation of linguistic awareness and vocabulary to word reading and spelling for first-grade students participating in response to instruction. *Language, Speech, and Hearing Services in Schools, 44*(4), 337-347. https://doi.org/10.1044/0161-1461(2013/12-0013)

Landerl, K. & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics, 21*(2), 243–262. https://doi.org/10.1017/S0142716400002058

Larkin, R. F. & Snowling, M. J. (2008). Comparing phonological skills and spelling abilities in children with reading and language impairments. *International Journal of Language & Communication Disorders, 43*(1), 111–124. https://doi.org/10.1080/13682820601178584

Larkin, R. F., Williams, G. J., & Blaggan, S. (2013). Delay or deficit? Spelling processes in children with specific language impairment. *Journal of Communication Disorders, 46*(5–6), 401–412. https://doi.org/10.1016/j.jcomdis.2013.07.003

Lervåg, A. & Hulme, C. (2010). Predicting the growth of early spelling skills: Are there heterogeneous developmental trajectories? *Scientific Studies of Reading, 14*(6), 485–513. https://doi.org/10.1080/10888431003623488

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*(1), 1-14. https://doi.org/10.1007/s11881-003-0001-9

Mackie, C. J., Dockrell, J., & Lindsay, G. (2013). An evaluation of the written texts of children with SLI: The contributions of oral language, reading and phonological short-term memory. *Reading and Writing, 26*(6), 865–888. https://doi.org/10.1007/ s11145-012-9396-1

McArthur, G. M, Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the “specifics” of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry, 41*(7), 869-874. https://doi.org/10.1111/1469-7610.00674

McCarthy, J. H., Hogan, T. P., & Catts, H. W. (2012). Is weak oral language associated with poor spelling in school-age children with specific language impairment, dyslexia or both? *Clinical Linguistics & Phonetics, 26*(9), 791–805. https://doi.org/10.3109/02699206.2012.702185

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Streiftau, S., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Démonet, J.-F., Longeras, E., Valdois, S., George, F., . . . Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction, 29*, 65–77. https://doi.org/10.1016/j.learninstruc.2013.09.003

Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN is not a measure of orthographic processing. Evidence from the asymmetric German orthography. *Scientific Studies of Reading, 13*(1), 1–25. https://doi.org/10.1080/10888430802631684

Netherlands Institute for Social Research (NISR) (2017). *Status scores.* https://www.scp.nl/Onderzoek/Lopend_onderzoek/A_Z_alle_lopende_onderzoeken-/Statusscores).

Norton, E. S. & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63*, 427-452. https://doi.org/10.1146/annurev-psych-120710-100431

NVLF (2016). *Richtlijn Logopedie bij Taalontwikkelingsstoornissen.* https://www.nvlf.nl/wp-content/uploads/sites/2/2020/10/Richtlijn-TOS.pdf

Ouellette, G. (2010). Orthographic learning in learning to spell: The roles of semantics and type of practice. *Journal of Experimental Child Psychology, 107*(1), 50-58. https://doi.org/10.1016/j.jecp.2010.04.009

Perfetti, C. A. & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-214). John Benjamins Publishing Company.

Protopapas, A., Fakou, A., Drakopoulou, S., Skaloumbakas, C. & Mouzaki, A. (2013). What do spelling errors tell us? Classification and analysis of errors made by Greek schoolchildren with and without dyslexia. *Reading and Writing, 26*(5), 615–646. https://doi.org/10.1007/s11145-012-9378-3

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. (2007). Writing through retellings: An exploratory study of language-impaired and dyslexic populations. *Reading and Writing, 20*(3), 251–272. https://doi.org/10.1007/s11145-006-9030-1

Rakhlin, N., Cardoso-Martins, C., Kornilov, S. A., & Grigorenko, E. L. (2013). Spelling well despite developmental language disorder: What makes it possible? *Annals of Dyslexia, 63*(3–4), 253–273. https://doi.org/10.1007/s11881-013-0084-x

Quémart, P. & Casalis, S. (2013). Visual processing of derivational morphology in children with developmental dyslexia: Insights from masked priming. *Applied Psycholinguistics, 36*(2), 345-376. https://doi.org/10.1017/S014271641300026X

Raven, J. & Raven, J. (2003). Raven progressive matrices. In R. S. McCallum (Ed.) *Handbook of nonverbal assessment* (pp. 223–237). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0153-4

Schittekatte, M., Groenvynck, H., Fontaine, J., & Dekker, P. (2007). *TEA-Ch: Test of Everyday Attention for Children, Dutch version.* Pearson.

Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL. Dutch version.* Harcourt Assessment.

Scuccimarra, G., Cutolo, L., Fiorillo, P., Lembo, C., Pirone, T., & Cossu, G. (2008). Is there a distinct form of developmental dyslexia in children with specific language impairment? Findings from an orthographically regular language. *Cognitive and Behavioral Neurology, 21*(4), 221–226. https://doi.org/10.1097/WNN.0b013e31818a5caf

SDN, De Jong, P.F., de Bree, E.H., Henneman, K., Kleijnen, R., Loykens, E.H.M., Rolak, M., … & Wijnen, F.N.K. (2016). *Dyslexie: Diagnostiek en behandeling: Brochure van de Stichting Dyslexie Nederland (SDN) [Dyslexia: diagnostics and treatment].* Retrieved from http://www.stichtingdyslexienederland.nl/publicaties/brochures-sdn

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Silliman, E. R., Bahr, R. H., & Peters, M. L. (2006). Spelling patterns in preadolescents with atypical language skills: Phonological, morphological, and orthographic factors. *Developmental Neuropsychology, 29*(1), 93–123. https://doi.org/10.1207/s15326942dn2901_6

Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., Hulme, C., & Wellcome Language and Reading Project Team. (2019). Developmental outcomes for children at high risk of dyslexia and children with developmental language disorder. *Child Development, 90*(5), e548-e564. https://doi.org/10.1111/cdev.13216

Treiman, R. & Kessler, B. (2014). *How children learn to write words.* Oxford University Press.

Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., & Lyytinen, H. (2010). Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *Journal of Learning Disabilities, 43*(4), 308–321. https://doi.org/10.1177/0022219410369096

van den Bos, K. P., & lutje Spelberg, H. C. (2007). *Continu Benoemen & Woorden Lezen (CB&WL) [Continuous Naming & Word Reading].* Amsterdam, The Netherlands: Boom testuitgevers.

van den Bos, K. P., lutje Spelberg, H. C., Scheepstra, A. J. M., & de Vries, J. R. (1994). *De Klepel. Vorm A en B [Klepel nonword reading].* Pearson.

Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2012). Auditory processing and speech perception in children with specific language impairment: Relations with oral language and literacy skills. *Research in Developmental Disabilities, 33*(2), 635–644. https://doi.org/10.1016/j.ridd.2011.11.005

van Viersen, S., de Bree, E. H., Verdam, M., Krikhaar, E., Maassen, B., van der Leij, A., & de Jong, P. F. (2017). Delayed early vocabulary development in children at family risk of dyslexia. *Journal of Speech, Language, and Hearing Research, 60*, 937–949. doi:10.1044/2016_ JSLHR-L-16-0031

van Weerdenburg, M., Verhoeven, L., Bosman, A., & van Balkom, H. (2011). Predicting word decoding and word spelling development in children with specific language impairment. *Journal of Communication Disorders, 44*(3), 392–411. https://doi.org/10.1016/j.jcomdis.2010.12.002

Werfel, K. L., Al Otaiba, S., Kim, Y.-S., & Wanzek, J. (2021). Linguistic predictors of single-word spelling in first-grade students with speech and/or language impairments. *Remedial and Special Education, 42*(2), 118-128 . https://doi.org/10.1177/0741932520918858

Windsor, J., Scott, C. M., & Street, C. K. (2000). Verb and noun morphology in the spoken and written language of children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*(6), 1322–1336. https://doi.org/10.1044/jslhr.4306.1322

Wolter, J.A., & Apel, K. (2010). Initial acquisition of mental graphemic representations in children with language impairment. *Journal of Speech, Language, and Hearing Research, 53*, 179–1

Supplemental figures

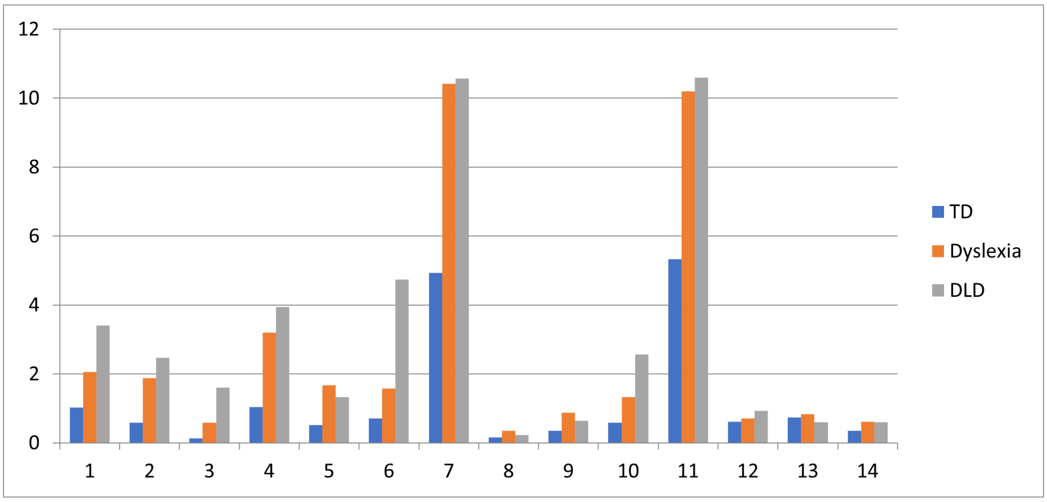


Figure 3: Number of Errors by SVS Spelling Category (de Vos & Braams, 2015, see Table 2) for Three Groups

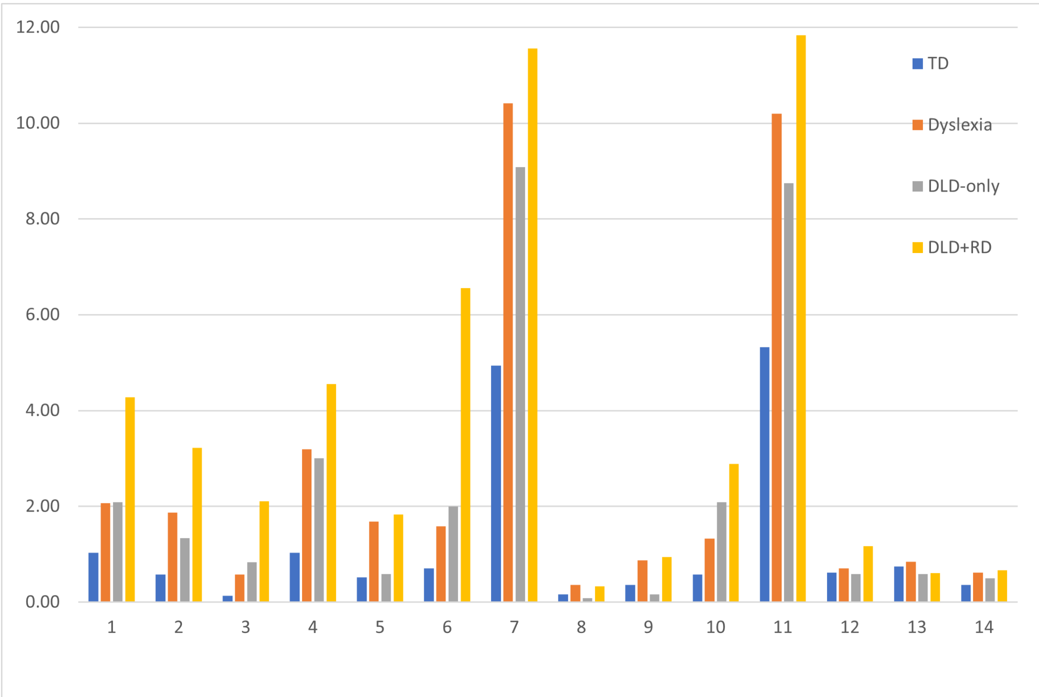


Figure 4: Number of Errors by SVS Spelling Category (de Vos & Braams, 2015, see Table 2) for Four Groups

STEM-, SPRAAK- EN TAALPATHOLOGIE

Gepubliceerd online: 5 november 2022

University of Groningen Press

Vol. 27, 2022, Special Issue, pp. 53-68
Special Issue for Steven Gillis

https://doi.org/10.21827/32.8310/2022-SG-53

Communication with diminutives to young children vs. pets in German, Italian, Lithuanian, Russian, and English

Wolfgang U. Dressler[1,2], Elisa Mattiello[3], Katharina Korecky-Kröll[1], Sabrina Noccetti[3], Ineta Dabašinskienė[4], Laura Kamandulytė-Merfeldienė[4], & Victoria V. Kazakovskaya[5]

[1] *University of Vienna*

[2] *Austrian Academy of Sciences*

[3] *University of Pisa*

[4] *Vytautas Magnus University, Kaunas*

[5] *Institute for Linguistic Sciences, Russian Academy of Sciences, St. Petersburg, Rusland*

Samenvatting

Dit artikel gaat over het gebruik van verkleinwoorden en koosnamen (hypocoristics) in twee taalregisters: taal gericht tot kinderen (child-directed speech, CDS) en taal gericht tot huisdieren (pet-directed speech, PDS). De semantiek van verkleinwoorden blijkt een minder grote rol te spelen dan de pragmatiek: de emotionele nabijheid van kinderen en huisdieren. De studie, waarin vijf talen worden vergeleken, verkent ook de typologie: de morfologische rijkdom van verkleinwoorden in een taal beïnvloedt de productie. Daarnaast speelt de semantische transparantie van verkleinwoorden crosslinguïstisch een rol. In CDS en PDS worden meer transparante verkleinwoorden gebruikt.

Abstract

This contribution is dedicated to Steven Gillis with whom we have collaborated since the nineties within the "Crosslinguistic Project on Pre- and Protomorphology in Language Acquisition" on both child speech (CS) and child-directed speech (CDS) and also about the development of diminutives (DIMs). We investigate parallels in the use of DIMs and of hypocoristics (HYPs) between CDS and pet-directed speech (PDS), whereas CS is only marginally dealt with. When relevant, also adult-directed speech (ADS), written or oral (especially from electronic corpora, wherever available) will be compared. The presuppositions of this investigation will be stated at the beginning of the Introduction (§ 1).

Correspondentieadres:
Wolfgang U. Dressler
University of Vienna
Department of Linguistics
Sensengasse 3a, 1090 Wien
E-mail: wolfgang.dressler@univie.ac.at

This involves several innovations (beyond descriptions of new data), when compared with existing literature, relevant to theoretical and typological problem areas.

We will show that also in DIMs and HYPs used in CDS and PDS semantics only plays a partial or even marginal role when using more DIMs to communicate with young children and young and/or small pets, because it is more relevant that both younger and smaller pets are emotionally closer to us, which is again a pragmatic factor.

In regard to language typology, we will apply our concepts of morphological richness and productivity, as argued for and supported in our previous publications, to CDS and PDS and show that richer and more productive patterns of DIM formation of a language also have a typological impact on more frequent and more productive use both in CDS and PDS.

We will also apply our concepts of grading morphosemantic transparency/opacity, as argued for and supported in our previous publications, and we start to show, as already shown for CS, that also in CDS towards young children (and similarly in PDS) more morphosemantically transparent DIMs are used than in ADS. This is also connected to their predominantly pragmatic meanings in CDS and PDS (obviously not exclusively pragmatic as in early CS).

The languages and authors were selected according to who among the participants in the Crosslinguistic Project on Pre- and Protomorphology in Language Acquisition had CDS and PDS available, plus Elisa Mattiello who has collected English and Italian PDS data.

INTRODUCTION

The presuppositions of our enterprise of comparing the use of diminutives (DIMs) and of hypocoristics (HYPs) formed with DIM suffixes in child-directed speech (CDS), pet-directed speech (PDS), also child speech (CS), are that

a) PDS has been extended diachronically from early child-centred speech, especially CDS, similarly to lover-centred speech, i.e., when lovers speak to or about each other, as in G. mein Lieb+i ‘my love/beloved+DIM’,

b) PDS represents a nearly totally asymmetric communication (in the sense of Watzlawick et al. 2011), whereas CDS considers the expectations of hearers much more (incl. non-addressees),

c) there is a predominance of pragmatics over semantics, especially in CS, where DIMs whose meanings are ascertained are at first only pragmatically used. This includes the impact of gender differences of speakers and addressed children vs. pets and the only exceptional relevance of semantics in the use of HYPs, which are two other pragmatic factors.

Moreover d) We also compare the impact (especially on CDS, CS and PDS) of the amount of morphological richness of language-specific DIM formation.

e) Finally, we study the impact of morphosemantic and morphotactic transparency on DIM usage in child-directed speech and pet-directed speech vs. adult-directed speech.

The data on CDS come from the collection of spontaneous interaction between children and their relevant caretakers within the Crosslinguistic Project on Pre- and Protomorphology in Language Acquisition and related language-specific projects, such as the Austrian INPUT Project. English data come from the Manchester corpus in the CHILDES database, collected by Theakston et al. (2001).

The PDS data come first of all from interviews on Italian, German, English and Russian performed for previous publications. The data were collected via questionnaires. The Lithuanian data were collected via questionnaires for the purposes of this study.

We investigate parallels in the use of diminutives (DIMs) between child-directed speech (CDS) and pet-directed speech (PDS). This involves several innovations when compared with existing literature, relevant to theoretical and typological problem areas.

A presupposition of this enterprise is (beyond the above-mentioned evaluation of PDS) that there is a predominance of pragmatics over semantics, especially in CS, where DIMs whose meanings are ascertained are first only pragmatically used. Moreover, the priority of pragmatics is evidenced by the favouring role of emotion, familiarity, and empathy and by DIMs being used more towards pets than about pets. It is dubious whether much irony or even sarcasm occurs in PDS and CDS (in contrast to humour).

The relevance of semantics is contradicted by PDS examples such as German:

‘So ein großes Hauf+i/+erl hast du gemacht!’
‘You produced such a big shit+DIM!’

In PDS, at least in part similar to CDS, semantics plays only a partial or even marginal role when using more DIMs to communicate with young children and young and/or small pets, because it is more relevant that both younger and smaller pets are emotionally closer to us, which is again a pragmatic factor. Also the interviewed native speakers of Russian and Lithuanian showed an analogous sensitivity, expressed in the major use of DIMs in relation to young and sick pets, as they reported in their response to the questionnaire.

As to language typology, the richer, more frequent and more productive DIM patterns are, the more DIMs are used both in CDS and PDS: Italian DIM formation is the richest and most productive, with the suffix -ino being the most frequent and productive suffix, also in CDS and PDS. Lithuanian comes next in its productive formation of DIMs from any noun via one or several competing suffixes. The suffixes (also used by the participants) are masc. *-elis/-ėlis, -(i)ukas, -utis, -ytis, -aitis*, fem. *-elė/-ėlė, -(i)ukė, -utė, -ytė, -aitė*.

Then comes the Russian language, whose number of DIM suffixes exceeds 30 (including allomorphs). The most frequent and productive among them in asymmetric communication are *-k, -ik, -očk, -ulj, -uš* as well as *-ok/ek* and *-en’k* in female PDS.

Then comes Viennese Austrian Standard German, where there is more preponderance of (Austrian) *-erl* over (common German) *-chen* in PDS than in CDS.

English DIM formation is the least rich, frequent in usage and productivity, both in CDS and PDS.

Finally, we will compare the impact of morphosemantic and morphotactic transparency on DIM usage in CDS and PDS, also referring to gender differences of the speakers and addressees.

History of Research

Our view on pragmatics is stated (with discussion of other views) in Dressler & Merlini Barbaresi (1994) and Merlini Barbaresi & Dressler (2020), who have supported in many publications on morphopragmatics their view that DIMs have primarily a pragmatic and not a semantic meaning, i.e., determined by speech situation and/or speech act. That this is even more the case in CS has been shown in many of our publications from Savickienė & Dressler (2007) up to Dressler et al. (2019).

Research on CDS, pioneered by Ferguson (1977), Snow (1977), Newport et al. (1977), has been done for a long time and the most relevant studies are cited in our publications Savickienė & Dressler (2007) and Dressler, Ketrez & Kilani-Schoch (2017). All these studies show that CS development is very much influenced by CDS, because children do not acquire adult target languages as they are represented by grammars, dictionaries and adult electronic corpora, but as they are realized in CDS. In their recent interdisciplinary literature review, Schick et al. (2022) compare research on CDS to communication of great apes and other vocally learning animals.

All chapters of Savickienė & Dressler (2007) show that the first ascertainable meanings of DIMs in CS are pragmatic, whereas semantic meanings of smallness (in its polysemy) are expressed via adjectives meaning SMALL. This is apparently also preponderantly the case for CDS. But this has not yet been systematically investigated. The most systematic research on DIMs in CDS is represented by Savickienė & Dressler (2007), Dressler & Korecky-Kröll (2015).

HYPs are mostly used for pragmatic reasons, but nearly never referring only to young age, and thus for the whole life-time of a person, a truly semantic meaning is exceptional. The same holds for PDS (see Mattiello et al. 2021).

Systematic research on PDS is so far limited to Mattiello et al. (2021). In Russian linguistics, especially dealing with the problems of colloquial speech, there is a tradition of studying PDS, cf. Ermakova (1988, 1998), Sirotinina (1999), Kitajgorodskaya & Rozanova (1999), Bajkulova (2008) and Ermolova (2015). In particular, the abundance of DIMs in PDS is noted in a study based on modern corpus data of oral speech (Ermolova 2015: 61). DIMs are found both in appeals to pets and in etiquette formulas addressed to them, as well as in the names of surrounding objects. Such DIMs are considered by Baikulova (2008: 15) to be lexical means of heightened emotionality in communication with animals. DIMs are included in the "range of endearments" in conversations with pets (Ermakova 1988: 245), on the one hand, and in "the passive vocabulary of dogs", i.e. in the comprehension of pets (Ermakova 1998: 96), on the other.

HYPs with and without DIM suffixes have been most systematically described by Merlini

Barbaresi (2001) for English. The use of HYPs has been shown to depend nearly always on pragmatic variables, and very rarely also on semantics.

The concept of morphological richness has been worked on since Dressler (1999), and its importance for language acquisition has been investigated most systematically in Xanthos et al. (2011), Xanthos & Gillis (2010) and Dressler et al. (2019): greater richness of a morphological system (in our case of DIM formation) facilitates acquisition.

Our view on morphosemantic transparency/opacity has been developed from Dressler (1985), to Dressler et al. (2016), Ransmayr et al. (2016) and Mattiello & Dressler (2019), cf. Talamo et al. (2016), resulting in models of grading morphosemantic transparency/opacity in various degrees of fineness.

Data

As to CDS, the Austrian German, Italian, Lithuanian and Russian electronic data stem from corpora in spontaneous interaction with young children collected for Savickienė & Dressler (2007), for German also in Dressler & Korecky-Kröll (2015), based on the Viennese INPUT project and its successors. The English data have been collected from native speakers as well as from previous research on the -y/ie suffix (mainly Merlini Barbaresi 2001). Additional Italian data on DIMs and HYPs come from a collection of examples provided by interviewed native speakers, compared with studies on Italian HYPs (e.g., Thornton 2004). As to Lithuanian and Russian, additional data were collected within the Crosslinguistic Project on Pre- and Protomorphology in Language Acquisition.

For PDS, the German, Italian and English data were collected by written and oral interviews for our publication Mattiello et al. (2021). In addition, we analyzed PDS data of eight German-speaking parents from the INPUT project in comparison to their CDS and ADS.

Lithuanian data were collected via questionnaires from 73 native Lithuanian speakers (65 female, 8 male). The participants were nearly all present or previous pet owners, mostly dogs and cats, some of them had parrots, fish, rabbits. Only four of the participants reported that they do not use DIMs towards or about pets.

Russian data (516 DIMs and HYPs) have been obtained from semi-formal oral or written interviews with 130 native speakers (102 females, 28 males, aged 14 to 69 years) in accordance with the questionnaire developed by Mattiello et al. (2021). The age groups were balanced in number, except for those in their twenties. The participants were all pet owners, mostly of cats (49%) and dogs (23%). 15% had both cats and dogs. The percentage of those having birds (mostly parrots, but also peacocks), cats and birds as well as cats, dogs and birds, is from 1% to 2% for each group. The owners of other pets (such as fishes, hamsters, polecats, turtles, rabbits, chinchillas, flying foxes and rats) were rare, less than 1% in each group. They were asked about their pets' names, nicknames and HYPs that they used when addressing the pets or speaking about them with either other family members or relatives and friends. Questions were also related to the use of DIMs for the pets' body parts, foods and drinks, objects and toys, the place where they sleep or eat, their excrements or vomits and activities. The first results were presented in Vienna in February 2022 at the annual meeting of the Crosslinguistic Project on Pre- and Protomorphology in Language Acquisition and are partially published in Kazakovskaya (2022).

For ADS, the German data nearly exclusively come from the exhaustive Austrian Media Corpus (AMC). Another source are the parental interviews conducted with the eight parents from the INPUT project. As for Italian, we compared our data with the Italian Web 2020 corpus (itTenTen20), made up of texts collected from the Internet and containing 12.4 billion words (end of December 2020). The data on British English were compared with those included in the much bigger English Web 2020 corpus (enTenTen20). The most recent version of the enTenTen corpus consists of 38 billion words. The texts were downloaded between 2019 and 2021. For Lithuanian ADS, the Corpus of Spoken Lithuanian was used.

Methodology

The Austrian German (cf. also Dressler & Korecky-Kröll 2015), Italian, Lithuanian, Russian and English electronic CDS data were analysed according to the world-wide accepted and used conventions of CLAN (MacWhinney 2000) and following the methodology described in Savickienė & Dressler (2007), which consists in finding verbal and non-verbal contextual evidence for pragmatic vs. semantic meaning of DIMs.

Our PDS data were analyzed according to established methods in interview analyses (cf. Soeffner & Hitzler 1994 and Lenz, Ahlers & Werner 2014). For the methodology of interview interpretation in general, cf. also Alvesson (2010). The main element is a group discussion of how to interpret specific text elements, where members of the group try to falsify in a Popperian way claims of interpretation: the interpretation that survives falsification attempts best, is taken up as the final interpretation. For ADS, the data from the exhaustive Austrian Media Corpus (AMC) have been systematically enriched with the methods of the Austrian Centre for Digital Humanities of the Austrian Academy of Sciences followed also here (cf. Ransmayr et al. 2016, Ransmayr 2018).

The Corpus of Spoken Lithuanian, developed at Vytautas Magnus University, Kaunas (see Dabašinskienė, Kamandulytė 2009, Kamandulytė-Merfeldienė, Balčiūnienė 2016, Kamandulytė-Merfeldienė 2017) has been morphologically annotated for the automatic linguistic analysis using the CHAT (Codes for the Human Analysis of Transcripts) software. For Lithuanian and Russian the methods of the electronic corpora (mentioned in § 4) were used, for Russian systematically enriched with the methods of several leading Institutes of the Russian Academy of Sciences (cf. Natsional'nyj korpus russkogo jazyka [The Russian National Corpus] 2005, Plungian 2009).

Results

Lithuanian is after Dutch the second-richest European language in DIM use. Lithuanian research on CS, CDS, ADS, and PDS confirms this fact. For example, only 4 of 73 participants, who took part in this study, reported that they do not use DIMs speaking towards or about pets. HYPs and DIMs are mostly used for pragmatic reasons in CS, CDS, and ADS. The same holds for PDS (cf. Mattiello et al. 2021). PDS examples are masc. *katin+ėlis* 'cat-DIM', *angel+ėlis* 'angel-DIM', *auks+iukas* 'gold-DIM', *maž+iukas* 'small-DIM', *Merf+iukas* (Merfis-DIM), fem. *šaunuol+ėlė* 'fine-DIM', *ger+utė* 'good-DIM', *bit+utė* 'bee-DIM', *šun+ytė* 'dog-DIM', *bulv+ytė* 'potato-DIM', *Bel+utė* 'Bela-DIM'. The most frequently used suffixes are: -*(i)ukas*, -*utė*.

DIM suffixes are used very frequently when talking towards or about pets. Even two main masc. cat names had a DIM suffix: *Snieg+utis* 'snow-DIM' and *Debes+ėlis* 'cloud-DIM'.

As to Lithuanian HYPs with DIM suffixes in PDS, feminine bases end in -*a*, masculine bases in -*as* or -*is*: *Leta* → *Let+utė*, *Kapsis* → *Kaps+iukas*, *Kaps+ytis*, cf. DIM names in CDS and ADS: *Rūta* → *Rūt+ytė*, *Rūt+elė*, *Saulius* → *Saul+iukas*, *Monika* → *Monik+utė*, *Elijus* → *Elij+ukas*, *Tomas* → *Tom+ytis*, *Tom+ukas*.

The following examples of CDS and ADS examples from our other languages are limited to HYPs which are formed with DIM suffixes (as already above for Lithuanian). CDS examples are:

Germani: *Paul* → *Paul+i*, *Peter* → *Peter+l*

English: *Linda* → *Linn+ie*, *Minerva*, *Mary*, *Marie*, etc. → *Minn+ie*, *Cynthia* → *Cind+y*, *Floyd* → *Floyd+ee*

Italian: masc. *Davide* → *David+ino*, *Lorenzo* → *Lorenz+ino*, *Jacopo* → *Jacop+ino*, *Ettore* → *Ettor+ino*, *Andrea* → *Andre+uccio* (above all in Tuscany); fem. *Chiara* → *Chiar+ina*, *Carlotta* → *Ott+ina*, *Sara* → *Sar+etta*, *Cinzia* → *Cinzi+etta*, *Viola* → *Viol+etta*. The suffix -*i* (borrowed from English) occurs in *Tommaso* → *Tomm+i/+y*, fem. *Michelle* and masc. *Michele* → *Mich+i/-y*, *Angelica* → *Ang+i/+y*.

Russian: masc. *Kirill* → *Kirj+uša*, *Slava* (from *Vjačeslav*) → *Slav+očka* / *+on'ka* / *+uška* / *+ulja* , *Filipp* → *Filipp+uša* / *+ok* / *+čik*, *Filja* → *Fil+en'ka*; fem. *Liza* (from *Elizaveta*) → *Liz+očka* / *+on'ka* / *+ok*, *Toma* (from *Tamara*)→ *Tom+očka* / *+usja* / *+čik*. In suffixation, more than one suffix may be used, as in *Fillip-oč+ek*, *Slav-ul+ečka*, *Tom-us+en'ka*.

Female gender is another pragmatic factor for favouring the use of HYPs (with DIM suffixes) of the speaker and/or of the addressee of both CDS and CS favours, since it is not the objective gender, but the identification of being of female gender, which is the decisive factor, see in Savickienė & Dressler (2007) for German (p.212-214) and in Kazakovskaya & Argus (2021: 87) for Russian and Estonian, and in Dabašinskienė (2012) for Lithuanian.

For Russian PDS, the more frequent use of DIMs and HYPs by female respondents of different ages was shown in Kazakovskaya (2022). Women are more likely than men used these words in all the situations mentioned above (§ 4), that is, talking to or about pets, their body parts, their food and drink, objects and toys, and so on (85% DIM and HYP in female PDS vs

15% - in male PDS, in types). However, these groups had different frequency in the speech of the respondents. Female speakers use from 4 to 24 DIM suffixes, whereas males use from 2 to 16. The most numerous and diverse suffixes are used by females when addressing pets, and when not calling them by name or when calling themselves pet-owners, similar to G. *dein Frau+erl* / *Herr+l* 'your mistress+DIM/master+DIM'.

Lithuanian participants reported that they use DIMs when the pet does something good (e.g., *auks+elis* 'gold-DIM', *ger+utis* 'good-DIM', fem. *maž+iukė* 'small-DIM', but even when the pet does something bad (e.g., *veln+iukas* 'devil-DIM', *blog+iukas* 'bad-DIM', *parš+elis* 'piglet-DIM').

One participant used for the pet *Dora* by assonance to its DIM-suffixed form *Dor+ytė* the neologistically created HYPic *Doryb+ytė* 'virtue, morals-DIM', sometimes with empathy, sometimes ironically.

In contrast to other languages, in Lithuanian PDS, the ironical usage of DIMs and HYPs is frequent, e.g., masc. *durn+elis* 'stupid-DIM', *smirdž+iukas* 'stinky-DIM', *storul+iukas* 'thick-DIM', *cepelin+ukas* 'dish zeppelin-DIM', *pabais+ulis* 'monster-DIM', fem. *vėmal+iuk+ininkė* 'vomit-DIM', *kvail+elė* 'stupid-DIM', *karv+ytė* 'cow-DIM'.

If we compare the CDS, PDS and ADS data of the same eight Austrian parents, we find the highest prevalence of HYPs over DIMs in PDS (77.27 % HYP vs. 22.73 % DIM tokens), followed by CDS (61.48 % HYP vs. 38.52 % DIM), whereas ADS uses clearly more DIMs than HYPs (71.88 % DIM vs. 28.13 % HYP). Thus, the pragmatic priority is also reflected in the high frequencies of HYPs in PDS and CDS. However, it must be noted that the rates of DIM tokens including HYPs among all noun tokens are also very different in the three settings: highest in PDS (22.22 %), followed by CDS (9.24%) and particularly low in ADS (1.70 %).

In Russian PDS, the proportion of DIMs is slightly higher than that of HYPs (44.6% HYP vs. 55.4% DIM), while in CDS DIMs significantly predominate (5.2% HYP vs. 94.8% DIM).

Typological results on the impact of morphological richness of diminutive formation

Richness of productive morphological patterns refers to the amount of productivity of morphological patterns, only secondarily to their frequency (cf. also Bauer's 2001 notion of profitability), because the frequent use of a morphological pattern may be only a fossilized residue of earlier productivity (see Dressler 1999, 2003).

In Dressler (1999, 2003), the author has also stressed the importance for language acquisition, insofar as greater morphological richness results in faster speed of acquisition of morphological patterns, as proved for inflection (Xanthos et al. 2011), for compounding (Dressler, Ketrez & Kilani-Schoch 2017), and probably for DIMs (as shown here below) by a reanalysis of the results in Savickienė & Dressler (2007) and of our other publications. Here we can refer, for example, to Dressler et al. (2021) for massive supporting data in the internet and child language and to Sommer-Lolei et al. (2021) for psycholinguistic experiments. One important consequence of morphological richness is the productivity of a pattern. This can be ascertained most easily in the applications to recently borrowed words.

For Italian we have established this (Dressler et al. 2019) for DIMs by finding on the internet thousands or at least hundreds of tokens of recent English loan-words, such as *weekend+ino* / *+etto* / *+uccio* and 30 similar sets of DIMs, whereas for German we found only two examples: *Computer+l* / *+chen*, *Roboter+l* / *+chen*. The wealth and productivity of Italian DIM formation is increased by interfixation (Dressler & Merlini Barbaresi 1992) in DIM formation of recent English borrowings, such as of -*er*- (e.g., *flopp+er+ello*), of -*ar*- (*scoop+ar+ello*) and by allomorphic insertion of -*c*-, as in *padron-c+ino* vs. *padron+ello*.

Lithuanian examples are *biudžet+ėlis*, *biudžet+ukas*, *reform+ėlė*, *pinkod+ukas*, *smail +iukas*, *fail+iukas*, *ok+iukas* 'OK-DIM', *hamburger+iukas*, *remiks+iukas*, *procent +ėlis* / *+ukas* 'percent-DIM' (thus, presumably borrowed earlier from G. Prozent).

In Russian ADS, DIMs from recent borrowings show the high productivity for the following suffixes: -*ik* as in *smajl+ik*, *prajs+ik*, *fail+ik*, *kompjut+ik* 'computer-DIM', *nout+ik* 'notebook-DIM', *mejl+ik*, *futb+ik* 'football-DIM', *koronavirus+ik*, *kovid+ik*, *kuar+ik* 'QR-DIM', -*ušk* in *kovid+uška*, *koron+uška* 'corona-DIM', *menj+uška* 'menu-DIM', *ferrar'+ka* 'ferrari-DIM', -*čik* in *ajfon+čik* 'I-phone-DIM', *miniven+čik* 'minivan-DIM', -*ok* / *ek* in *xot-dož+ek* 'hot-+dog-DIM', as well as -*očk* in *koronar+očka* 'corona-DIM', -*išk* in *juesb+iška* 'USB-DIM'.

For English this criterion is not applicable, because there are too few recent borrowings from other languages. Moreover, we doubt that DIMs could be derived from relatively recent German borrowings such as Weltanschauung, Blitzkrieg.

In German CDS the suffixes -*i* (most frequent, e.g., *Mam+i* 'mummy'), -*erl*, -*chen*, -*li*, unproductive -*lein* (Korecky-Kröll & Dressler 2007) occur, e.g., in *Herz+i* 'heart-DIM', *Eng+erl* 'angel-DIM', *Schwester+chen* 'sister-DIM', *Ohr+li* 'ear-DIM', *Maus+i+lein* 'mousie'.

The Italian suffixes are in order of frequency of use -*ino* > -*etto* > -*ett+ino*, -*otto*, -*ello*, -*ol*-*ino*, plus feminines in -*a*, as in *can+ino* 'dog-DIM', *capr+etta* 'goat-DIM', *ors+ett+ino* 'bear-DIM-DIM', *bambol+otto* 'doll-DIM', *asin+ello* 'donkey-DIM', *conigli+ol+ino* 'rabbit-DIM-DIM'.

The most productive Lithuanian suffixes are masc. -*elis* (*nam+elis* 'house-DIM'), -*ukas* (*nam+ukas* 'house-DIM') and fem. -*elė* (*saul-elė* 'sun-DIM'), -*ytė* (*saul-ytė* 'sun-DIM').

The most frequent Russian suffixes are -*k* (*golov+ka* 'head-DIM', *obezjan+ka* 'monkey-DIM), -ik (*samoljot+ik* 'airplane-DIM', *šar+ik* 'ball-DIM'), -*ok*/*ek* (*čaj+ok* 'tea-DIM', *oreš+ek* 'nut-DIM').

In ADS, English examples are limited to a single suffix: *hubb+y* 'husband', *siss+y* 'sister', *shopp+ie* 'shop-assistant, *best+ie* 'best friend'. For our other languages, see the DIM suffixes listed above.

The mass of Lithuanian PDS examples and the relatively frequent ironic use of DIMs and of HYPs is another evidence for Lithuanian being, after Dutch, the DIM-richest European language in forms (langue) and uses (parole).

Results on morphosemantic transparency vs. opacity

Dressler et al. (2019) have found, as expected in the framework of Natural Morphology, that DIMs in CS are more often morphosemantically transparent derivations than in ADS. Presumably also CDS DIMs are more often transparently derived than ADS DIMs. But this has not yet been investigated empirically.

In Schwaiger et al. (2019), we investigated this question with respect to the more transparent colloquial -*erl* vs. the less transparent, often umlauted Standard German -*chen* DIMs. However, as speaking Standard German to their children seems to be more important for Viennese parents than talking in a (morphologically) transparent way, they use more opaque -*chen* DIMs (including German ones as spoken in Germany) in CDS than more transparent -*erl* DIMs. Nevertheless, in Schwaiger et al. (2019) we did not take into account the most transparent -*i* DIMs which have been shown to have high type frequencies and particularly high token frequencies in parents' speech to very young children (see Korecky-Kröll & Dressler 2007 and Korecky-Kröll 2011: 342-346) and which are particularly used to and by young children.

In the data of the eight Austrian parents, we find 100 % of transparent DIMs in PDS, 93.75 % of transparent DIMs in ADS and 92.01 % of transparent DIMs in CDS, which shows again that the use of Standard DIMs is more important than the use of exclusively transparent DIMs, at least in CDS with older kindergarten children aged 3 to 5 years. This is also true for Russian CDS and CS vs. ADS. In general, Russian DIMs are transparent due to the transparency of their DIM suffixes (although there are a number of alternations when a suffix is attached to a root). A problem arises with the suffix -*k* in case of loss of DIM meaning in many noun-derived DIMs, such as *kolen+ka* 'knee', *kartoš+ka* 'potato' or *seljod-ka* 'herring'. Russian grammar (e.g., Švedova 1980/2005) refers to them not as DIMs, but as stylistic-colloquial modifications of nouns. However, sometimes such words are considered as DIMs by relevant publications.

For PDS, Mattiello et al. (2021) have found that nearly all DIM derivations are morphosemantically transparent and pragmatically used, e.g. It. empathetically used *nas+ino* 'nose+DIM', *cod+ina* 'tail+DIM', *zamp+ina* 'paw+DIM'.

The same holds for DIMs in Russian PDS, as in *moloč+ko* 'milk-DIM', *kolbas+ka* 'sausage-DIM', *krovat+ka* 'bed-DIM'; *kljuv+ik* 'beak-DIM', *us+iki* (Plural) 'mustashe-DIM', *dom+ik* 'house-DIM', *pled+ik* 'plaide-DIM'; *jazyč+ok* 'tongue-DIM', *syn+ok* 'son-DIM'.

The default suffixes for DIM formation in Standard Lithuanian are masc. -*elis* and fem. -*elė* (Ambrazas 1997). They cover a wide range of pragmatic functions, but also encompass the meaning of smallness.

However, the CDS data do not conform with this pattern, as other suffixes, more often associated with the meaning of "small" and "young", -*ukas* (masc., 60%) and -*ytė* (fem., 50%) show the highest frequency (Savickienė 2007). The results of PDS support this tendency: the same suffixes are reported to be used more commonly when addressing or talking about the pet.

Moreover, DIM suffixes are more often morphosemantically transparent in CDS and PDS than in ADS, and, as expected, used for pragmatic functions. PDS examples are masc. *šun*

+(i)ukas 'dog-DIM', *nos+(i)ukas* 'nose-DIM'; fem. *dešr+ytė* 'sausage-DIM', *mės+ytė* 'meat-DIM', *galv+ytė* 'head-DIM'.

Conclusions and outlook

Our previous claims on the priority of pragmatics over semantics in evaluative morphology have been supported by the nearly exclusive pragmatic use of DIMs in PDS and by the much more frequent pragmatic than semantic motivation of the use of HYPs with DIM suffixes beyond the mostly conventional use of such HYPs for children in CDS.

The impact of the morphological richness of DIM formation on the productive use of DIMs in ADS and CDS has been supported in more languages than in our previous publications. We have made progress in various aspects of the comparative study of CDS

We have obtained some data which show differences in PDS according to the age of the speakers.

How profitable our investigations are, shows also in the following outlook on which lacunae still have to be filled: In a future paper planned to partially overlap with the topics of this paper, we will use formal semantic results of Flaig's (2022) forthcoming MA thesis (supervised by the first author in Vienna) for differentiating semantic and pragmatic meanings of DIMs, incl. the relations between CDS and CS, and for the scaling of semantic transparency/opacity of German and Italian DIMs.

In addition, we plan a systematic investigation showing that also in CDS the first ascertainable meanings of DIMs are preponderantly pragmatic, whereas semantic meanings of smallness (in its polysemy) are mainly expressed via adjectives meaning SMALL, albeit to a lesser extent than in CS, and whether the chronological development shows some fine-tuning. An analogous lacuna should be filled for morphosemantically transparent vs. opaque DIM derivations (see § 8). In general, systematic investigation of the development of CDS and CS is needed for searching evidence of fine-tuning in the domain of DIMs and HYPs.

What also needs a systematic investigation is whether many of the Italian HYPs with the suffix -[i] used in CDS vs. PDS vs. CS vs. ADS are simply borrowed from English or are formed in Italian with the borrowed suffix -y. A further lacuna is which patterns of the formation of HYPs via DIM suffixations are still productive.

References

Alvesson, M. (2010). *Interpreting Interviews.* London: Sage.

Ambrazas, V. (1997). Ed. *Dabartinės lietuvių kalbos gramatika (The grammar of modern Lithuanian).* Vilnius: Mokslo ir enciklopedijų leidykla.

Bajkulova, A.N. (2008). Rečevoe obščenie v sem'je: rol' v njom životnyx, rastenij i predmetov veščnogo mira (Speech Communication in a Family: the Role of Pets, Plants and Inanimate Objects in it). Izvestija Saratovskogo universiteta. Ser. Filologija. *Žurnalistika 8* (2), 14-20.

Bauer, L. (2001). *Morphological Productivity.* Cambridge: Cambridge University Press.

Dabašinskienė, I. (2012). Gender differences in language acquisition : a case study of Lithuanian diminutives. *Journal of Baltic studies 43*, 177-196.

Dabašinskienė I., Kamandulytė L. (2009). Corpora of Spoken Lithuanian. *Estonian Papers in Applied Linguistics 5*, 67-77.

Dressler, W.U. (1985). On the predictiveness of Natural Morphology. *Journal of Linguistics 21*, 321-337.

Dressler, W.U. (1999). Ricchezza e complessità morfologica. In P. Benincà, A. Mioni, & L. Vanelli (Eds.), *Fonologia e Morfologia dell'Italiano e dei Dialetti d'Italia. Atti del 21. congresso SLI.* Roma: Bulzoni, 587-597.

Dressler, W. U. (2003). Degrees of grammatical productivity in inflectional morphology. *Italian Journal of Linguistics 15*, 31–62.

Dressler, W.U. & Merlini Barbaresi, L. (1992). How to fix interfixes? On the structure and pragmatics of Italian (and Spanish, Russian, Polish) antesuffixal interfixes and of English "intermorphic elements". *Acta Linguistica Hungarica 36*, 53-67.

Dressler, W.U. & Merlini Barbaresi, L. (1994). *Morphopragmatics.* Berlin: Mouton de Gruyter.

Dressler, W.U., Ketrez, F.N., & Kilani-Schoch, M. (2017). Eds. *Nominal Compound Acquisition.* Amsterdam: Benjamins.

Dressler, W.U. & Korecky-Kröll, K. (2015). Evaluative morphology and language acquisition. in N. Grandi & L. Körtvélyessy (Eds.) *Edinburgh Handbook of Evaluative Morphology.* Edinburgh: Edinburgh University Press, 134-141.

Dressler, W.U., Korecky-Kröll, K., Christofidou, A. & Mörth, K. (2016). Genus und Deklination im Deutschen und Griechischen im System und im Erstspracherwerb: eine kontrastive Untersuchung (Gender and declension in German and Greek in system and language acquisition: a contrastive investigation). in A. Bittner & C. Spieß (Eds.) *Formen und Funktionen: Morphosemantik und grammatische Konstruktion.* Berlin: de Gruyter, 61-84.

Dressler, W.U., Merlini Barbaresi, L., Schwaiger, S., Ransmayr, J., Sommer-Lolei, S. & Korecky-Kröll, K. (2019). Rivalry and Lack of Blocking Among Italian and German Diminutives in Adult and Child Language. in F. Gardani, W.U. Dressler & H.C. Luschützky (Eds.) *Competition in Inflection and Word-Formation.* Cham: Springer Nature Switzerland, 123-143.

Ermakova, O. P. 1988. Razgovory s životnymi (lingvo-psixologičeskie zametki) (Conversations with Animals (Linguistic and Psychological Notes)). in D.N. Šmeljov & E.A. Zemskaja (Eds.) *Raznovidnosti gorodskoj ustnoj reči. Sbornik naučnyx statej.* Moscow: Nauka, 240-247.

Ermakova, O.P. (1998). Rečevoj portret sobaki (The Speech Portrait of a Dog). in M.J. Glovinskaja (Ed.) *Liki jazyka. K 45-letiju naučnoj dejatel'nosti E. A. Zemskoj.* Moscow: Nasledie Pub., 94-104.

Ermolova, O.P. (2015). Lingvističeskie osobennosti obščenija čeloveka s domašnimi životnymi (Linguistic features of human-to-pet communication). Vestnik Permskogo universiteta. *Rossijskaja i zarubežnaja filologija 4*(32), 58–66.

Flaig, J. (2022). *Application and comparison of new theoretical approaches in semantics* *(particularly of diminutives) based on the NSM approach* by Anna Wierzbicka. Dept. of Linguistics, Vienna University: forthcoming MA thesis.

Kamandulytė-Merfeldienė, L. (2017). Sakytinės lietuvių kalbos tekstynas – natūralios vartosenos tyrimų šaltinis (The Corpus of Spoken Lithuanian as a Research Source of Natural Usage). *Taikomoji kalbotyra 9*, 176-198.

Kamandulytė-Merfeldienė, L. & Balčiunienė, I. (2016). Syntactically coded corpus of spoken Lithuanian: developmental issues and pilot studies. *Kalbų studijos 28*, 92–101.

Kazakovskaya, V.V. (2022). Diminutivy v reči, obraščennoj k detjam i k domašnim životnym (Diminutives in child- and pet-directed speech). in M. Elivanova et al. eds. *Problemy ontolingvistiki – 2022: rečevoj mir rebjonka (universal'nye mexanizmy i individual'nye processy.* St. Peterburg: Izdatel'stvo VVM, 22—27.

Kazakovskaya, V.V. & Argus, R. (2021). Acquisition of diminutives in typologically different languages: evidence from Russian and Estonian. *Estonian Papers in Applied Linguistics 15*, 81–97. doi:10.5128/ERYa17.05

Kitajgorodskaja, M.V. & Rozanova, N.N.. 1999. *Reč moskvičej. Kommunikativno - kul'turologičeskij aspekt (The Speech of Moscow Citizens. Communicative and Cultural Approach).* Moscow: Russkie slovari.

Korecky-Kröll, K. (2011). *Der Erwerb der Nominalmorphologie bei zwei Wiener Kindern: Eine Untersuchung im Rahmen der Natürlichkeitstheorie.* University of Vienna, Dept. of Linguistics: PhD thesis.

Korecky-Kröll, K. & Dressler, W.U. (2007). Diminutives and Hypocoristics in Austrian German (AG). in I.Savickiene & W.U. Dressler (Eds.) *The Acquisition of Diminutives.* Amsterdam: Benjamins, 207-230.

Lenz, A., Ahlers, T., & Werner, M. (2014). Zur Dynamik bairischer Dialektsyntax – eine Pilotstudie. *Zeitschrift für Dialektologie und Linguistik 81*, 1–33.

Merlini Barbaresi, L. 2001. The pragmatics of the 'diminutive' -y/ie suffix in English. in C. Schaner-Wolles, J. Rennison, & F. Neubarth (Eds.) *Naturally! Linguistic Studies in Honour of Wolfgang Ulrich Dressler Presented on the Occasion of his 60th Birthday.* Torino: Rosenbach and Sellier, 315-326.

Merlini Barbaresi, L. & Dressler, W.U. 2020. Pragmatic Explanations in Morphology. in V. Pirrelli, I. Plag, & W.U. Dressler (Eds.) *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon.* Berlin: de Gruyter, 405-451.

Mattiello, E. & Dressler, W.U. (2019). The Morphosemantic Transparency/Opacity of Novel English Analogical Compounds and Compound Families. *Studia Anglica Posnaniensia 53*, 67–114. doi:10.2478/stap-2018-0004.

Mattiello, E., Ritt-Benmimoun, V., & Dressler, W.U. (2021). Asymmetric use of diminutives and hypocoristics to animals in Italian, Arabic, German and English. *Language & Communication 76*, 136-153.

Natsional'nyj korpus russkogo jazyka: 2003–2005. *Rezul'taty i perspektivy (The Russian National Corpus 2003—2005: results and perspectives).* (2005). Moscow: Indrik Publ.

Newport, E.L., Gleitman, H., & Gleitman, L.R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of motherese. in C. E. Snow & C.A. Ferguson (Eds.) *Talking to children: Language input and acquisition.* Cambridge: Cambridge University Press, 109–149.

Plungian, V.A. (2009). ed. *Natsional'nyj korpus russkogo jazyka: 2006–2008. Novye rezul'taty i perspektivy (The Russian National Corpus: 2006—2008. New results and perspectives).* St. Petersburg: Nestor-Istoria Publ.

Ransmayr, J. (2018). Austrian Media Corpus. in I. Börner, W. Soutlooktraub, & C. Zolles (Eds.) Germanistik digital. *Digital Humanities in der Sprach- und Literaturwissenschaft.* Wien: Facultas, 168-182.

Ransmayr, J., Schwaiger, S., Durco, M., Pirker, H., & Dressler, W.U. (2016). Graduierung der Transparenz von Diminutiven auf –chen: Eine korpuslinguistische Untersuchung (Gradation of transparency of diminutives in -chen). *Deutsche Sprache 44*, 261-286.

Savickienė, I. (2007). Form and meaning of diminutives in Lithuanian child language. In I. Savickienė & W.U. Dressler (Eds.) *The Acquisition of Diminutives: a cross-linguistic perspective.* Amsterdam: Benjamins, 13-41.

Savickienė, I. & Dressler, W.U. (2007). eds. *The Acquisition of Diminutives: a cross-linguistic perspective.* Amsterdam: Benjamins.

Schick, J., Fryns, C., Wegdell, F., Laporte, M., Zuberbühler, K., van Schaik, C.P., Townsend, S.W., & Stoll, S. (2022). The function and evolution of child-directed communication. *PLoS Biol 20*(5): e3001630. https://doi.org/10.1371/journal.pbio.3001630

Sirotinina, O. B. (1999). *Rečevoj portret sobaki: ill'uzija ili real'nost'? (The Speech Portrait of a Dog: An Illusion or Reality?)* Voprosy stilistiki: Mežvuzovskij sbornik naučnyx trudov 28. Saratov: Izdatel'stvo Saratovskogo universiteta, 167-174.

Snow, C.E. (1977). The development of conversation between mothers and babies. *Journal of Child Language 4*, 1–22. https://doi.org/10.1017/S0305000900000453.

Soeffner, H.-G. & Hitzler, R. (1994). Hermeneutik als Haltung und Handlung: über methodisch kontrolliertes Verstehen (Hermeneutics as attitude and action: on methodically controlled understanding). in N. Schröer (Ed.) *Interpretative Sozialforschung: auf dem Wege zu einer hermeneutischen Wissenssoziologie.* Opladen: Westdeutscher Verlag, 28-54.

Sommer-Lolei, S., Korecky-Kröll, K., Christiner, M., & Dressler, W.U. (2021). Acquisition and processing of (mor)phonotactic consonant clusters in German. in W.U. Dressler, B. Calderone, S. Sommer-Lolei & K. Korecky-Kröll (Eds.) *Experimental and Acquisitional Approaches to Morphonotactics.* Wien: Austrian Academy of Sciences Press, 77-100.

Švedova, N. J. (1980/2005). ed. *Russkaja grammatika* (Russian Grammar). Vol. 1. Moscow: Nauka Publ.

Talamo, L., Celata, C., & Bertinetto, P.M. (2016). DERIVATARIO: An annotated lexicon of Italian derivatives. *Word Structure 9*, 72-102.

Theakston, A.L., Lieven, E.V.M., Pine, J.M., & Rowland C. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language 28*, 127-152.

Thornton, A.M. 2004. Antroponimia. in M. Grossmann & F. Rainer (Eds.) *La formazione delle parole in italiano* (pp 601-607). Tübingen: Niemeyer.

Watzlawick, P., Bavelas, J.B., Jackson, D.D., 2011. *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes.* Norton, New York

Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., Gagarina, N., Hrzica, G., Ketrez, F.N., Kilani-Schoch, M., Korecky-Kröll, K., Kovačević, M., Laalo, K., Palmović, M., Pfeiler, B., Voeikova, M.D., & Dressler, W.U. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Language 31*, 461-479.

Appendix

Questionnaire for the pet study:

(M) or (F) AGE NATIONALITY

1) Do you have a pet animal?
If you don't, please enumerate other pets that come to your mind (cats, dogs, parrots or other birds, rabbits, guinea pigs, etc.)

2) If you do, how is it called? Do you also have a nickname or a hypocoristic name for it?

3) Do you use special diminutive forms when you address it? Please specify.

4) And when you speak about it (with friends, within the family)? e.g.
My pet is sick; or It has given birth to babies. Please specify if you use diminutive -y/ie forms.

5) Do you use special diminutive forms when you refer to its body parts?
a) in addressing the animal: Does anything hurt you, e.g. your mouth, your nose, snout, ear, belly, wings, paws, tail? Please specify.
b) in speaking about the animal: e.g.
My pet has broken its leg.
Its face is so sweet.
Please specify.

6) Do you use special diminutive forms when you refer to foods and drinks (animal feed, seeds, milk, water)? Please specify.

7) Do you use special diminutive forms when you refer to your pet's objects/toys (bed, ball, dog leash, cover, etc.)? Please specify.
Or when you refer to the place where it sleeps, eats (doghouse, kennel, litter, cage, stall, paddock, henhouse, fish tank, etc.)? Please specify.

8) Do you use special diminutive forms when you refer to your pet's excrements or vomit? Please specify.

9) Do you refer to yourself in your function as the pet's master or mistress? (Just for dogs): e.g., Your mistress/mommy is sad that you are sick. When returning after a long absence: Your mistress/mommy is finally again with you.

10) Can you refer to the pet's activities with a diminutive? Or can you form a phrase where a noun is put into a diminutive?

STEM-, SPRAAK- EN TAALPATHOLOGIE

Vol. 27, 2022, Special Issue, pp. 69-92
Special Issue for Steven Gillis

Gepubliceerd online: 5 november 2022

University of Groningen Press
https://doi.org/10.21827/32.8310/2022-SG-69

F0 dynamics associated with prominence realisation in children with hearing impairment

Jérémy Genette[1] & Jo Verhoeven[1,2]

[1] *Computational Linguistics & Psycholinguistics (CLiPS) Research Center, University of Antwerp*
[2] *Department of Language and Communication Science, City, University of London*

Samenvatting

Prominentie in spraak wordt typisch gerealiseerd door een grotere amplitude, duur en F0 van de klinkers in lettergrepen met zinsaccent/woordaccent. Het is algemeen bekend dat F0 de belangrijkste dimensie is. Deze studie doet verslag van de analyse van F0 in de uitspraak van woorden van twee groepen kinderen die verschillen in hoorstatus. De ene groep bestond uit kinderen met een gehoorbeperking, terwijl de andere groep bestond uit leeftijdsgenoten met een normaal gehoor. De kinderen met een gehoorbeperking hadden een cochleair implantaat of een conventioneel hoortoestel. De kinderen namen deel aan een woordimitatietaak die bestond uit het herhalen van monosyllabische woorden met een van de klinkers van het Belgisch Standaardnederlands. Uit de studie blijkt dat er interessante F0 verschillen tussen de groepen zijn. De kinderen met gehoorbeperking hebben de hoogste gemiddelde F0. Wat betreft de dynamiek van F0 realiseerden alle kinderen prominentie met een onderliggend stijgend-dalend patroon dat zich op fonetisch niveau manifesteerde als een dalende toonhoogtebeweging. Bovendien was de contour bij kinderen met een conventioneel hoortoestel het steilst, terwijl hij bij kinderen met een cochleair implantaat het vlakst was. De toonhoogtecontouren van kinderen met een normaal gehoor lag tussen de twee voorgaande groepen in. De waargenomen verschillen worden toegeschreven aan de aard van het hoortoestel.

Abstract

Prominence in speech is typically realised by means of greater amplitude, duration and F0 of the vowel nucleus in the syllable that carries word/sentence stress. It is well-established that F0 is the more important physical dimension. The present study reports the analysis of F0 in word realisation of two groups of children differing in hearing status. One group consisted of children with hearing impairment, while the other group consisted of age-matched children with normal hearing. The hearing-impaired children had been fitted with either a cochlear implant or a conventional hearing aid. Children had participated in a (non-)word imitation task which consisted of the repetition of

Correspondentieadres:
Jérémy Genette
University of Antwerp
Computational Linguistics and Psycholinguistics (CLiPS) Research Center
Prinsstraat 13, 2000 Antwerpen, Belgium
E-mail: jeremy.genette@uantwerpen.be

monosyllables containing one of the monophthongs of Belgian Standard Dutch. Measurement and analysis of F0 in the vowel nuclei revealed interesting differences between the groups. The children with hearing impairment had the highest overall F0. In terms of the dynamics of F0 associated with prominence, all children correctly realised an underlying prominence-lending rise-fall pattern which at the phonetic level manifested itself as a falling pitch movement. In addition, the contour in children with a conventional hearing aid was steepest, while it was shallowest in children with a cochlear implant. The contour in children with normal hearing was situated between the two previous groups. The observed differences are attributed to differences in device use.
Keywords: prominence, fundamental frequency, prosody, language acquisition, hearing impairment.

Introduction

Prosody in speech is very important to convey linguistic and paralinguistic information and to generally support efficient communication. Prosody is an umbrella term covering sentence-level phenomena such as intonation, pausing, rhythm, speech rate, voice quality and prominence. One of the functional roles of prosody in speech is to signal prominence. Terken & Hermes (2000) define prominence as a property of linguistic units which makes them stand out from their local context. Adequate use of prominence makes some parts of utterances more salient, typically those parts of utterances which are considered to convey new information in a communicative context. The phonetic characteristics of the Dutch intonation system have been particularly well described in ‘t Hart et al. (1990). This model of Dutch intonation describes intonation contours as sequences of perceptually relevant pitch movements which are defined phonetically in terms of a small number of acoustic parameters, i.e. their direction, timing, size and the rate of F0 change: some of these pitch movements are prominence-lending, while others are not. The specific acoustic characteristics of these pitch movements according ‘t Hart et al. (1990) are summarised in Table 1.

Prominence, in this model, is marked by a perceptually relevant pitch movement or a combination of pitch movements on the syllable which is required to stand out. The model incorporates two prominence-lending pitch movements: a rising and a falling one. The rising movement (1) is located very early in the vowel nucleus, it is abrupt and has a short duration of 100 msec. The falling movement (A) is located late in the vowel nucleus, it is abrupt and it has a short duration of 75 msec. Both pitch movements can be combined into 1&A pitch contours which also signal prominence. In other words, syllable prominence can be realised as a prominence-lending rising (1), falling (A) or as a combination of a rising and a falling (1&A) pitch movement within the same syllable nucleus. However, it should be noted that, in the phonological grammar of Dutch intonation, an A fall cannot occur without a preceding rise (’t Hart & Collier, 1979).

Table 1: Pitch movements taxonomy from 't Hart et al. (1990) and standard values for perceptually relevant pitch movements in Dutch (adapted from van Geel, 1983)[*: depends on accent interval; /: undefined].

Labels		1	2	3	4	5	A	B	C	D	E
Direction	Rise	X	X	X	X	X					
	Fall						X	X	X	X	X
Timing	Early	X				X		X			X
	Late			X			X				
	Very late		X						X		
Rate of change	Fast	X	X	X		X	X	X	X		X
	Slow				X					X	
Size	Full	X	X	X		X	X	X	X	X	
	Half				X						X
Duration (in ms)		100	100	150	*	50	75	75	20-50	*	38
Excursion (in ST)		+4	+4	+6	*	+2	-5	-5	/	*	-2.5
Slope (in ST/s)		40	40	40	40	40	66.666	66.666	/	*	65.789

Prosody and hearing impairment

According to the Universal Neonatal Hearing Screening (UNHS), the prevalence of hearing impairment among new-borns is 3 in 1000. The locus and severity of the impairment determine whether the rehabilitation should be carried out via a cochlear implant (henceforth: CI) or a conventional hearing aid (henceforth: HA) (Korver et al., 2017). Neither of these devices provide users with standard hearing but they improve hearing ability sufficiently to make communication possible.

Although it is now well established that the implantation of a CI in congenitally deaf children improves general language acquisition (e.g., Dettman et al., 2016; Levine et al., 2016), CI devices do not transmit all the intensity, temporal and spectral information present in the original acoustic signal (Moore, 2003; Green et al., 2004; Fu & Nogaki, 2005; O'Halpin, 2010). Furthermore, the access to auditory input for CI children is delayed up to the time of activation of the device. Similarly, the range of frequencies amplified by conventional hearing aids is typically limited to the range from 125 Hz to 8 kHz (Metcalfe, 2017). Although the distortions are less significant than those caused by CI, the provided input to HA is more restricted than in normally-hearing children. In other words, both HA and CI provide distorted input, but the extent of the distortions is higher for CI than HA.

Because of the imperfect access to the acoustic signal provided by these devices, hearing-impaired children's speech production differs from that of their NH peers both at the segmental level (cfr. Baudonck et al., 2010) and at the prosodic level (cfr. Lenden & Flipsen, 2007; Vanormelingen et al., 2016).

At the segmental level, the speech of hearing-impaired children is characterised by the distortion of both vowels and consonants. Common problems in the articulation of consonants involve voicing errors (voiceless sounds become voiced and vice versa) and place of articulation substitution errors typically associated with sounds that are articulated posteriorly in the oral cavity where articulatory gestures are less visible. In addition, consonant omission errors have been documented: in some studies word-initial consonant omission appears most frequently (Hudgins & Numbers, 1942), while in others consonant deletion was predominantly word-final (Nober, 1967; Markides, 1970; Smith, 1975). Furthermore, errors pertaining to consonant clusters have been noted: these often have to do with the omission of one of the consonants in clusters or by the insertion of schwas (e.g., Baudonck et al., 2010). The articulation of vowels also seems impaired, be it altogether less frequently than that of consonants. In children with a conventional aid several types of errors have been documented. Vowel substitutions are common and the findings suggest that back vowels are produced more correctly than front vowels and open vowels are more often correct than vowels with a closer degree of stricture (Smith, 1975; Geffner, 1980; Ozbič & Kogovšek, 2008, 2010). Nevertheless, the fronting of back vowels has also been reported (Stein, 1980). Another frequent error involves the neutralization of the peripheral vowels, i.e. the reduction of vowels to a more schwa-like quality (Markides, 1970; Smith, 1975). Furthermore, there have been reports of inappropriate vowel nasalization (Stevens et al., 1976) and the diphthongization of monophthongs (Markides, 1970; Smith, 1975).

At the prosodic level, inappropriate production of prosodic features has been described in hearing-aided speakers, but the findings are quite diverse. As far as the production of prominence is concerned, some studies describe deviant F0 contours (e.g. O'Halpin, 1997), while others mainly find durational differences (e.g. Nickerson, 1975). Sussman & Hernandes (1979) found that in the production of prominent syllables rising pitch contours were replaced by increased intensity. The absence of rising pitch movements on stressed syllables was also found by Gold (1980), Maassen (1986), and Osberger and McGarr (1982). In general, these studies concluded that hearing-aided speakers either produced excessive stress on prominent syllables or used flat and monotone pitch patterns throughout their utterances.

The main objective of the present paper is to study the precise characteristics of the F0 associated with word-level prominence in the speech of children with hearing impairment, as compared to their normally-hearing peers. More specifically, the purpose of this research is twofold. First, it provides a quantitative and empirical comparison of F0 production in children with and without hearing impairment by addressing the average F0 over the entire length of the vowel. Second, it focuses on the dynamics of F0 within the vowel itself.

Method

The data for this study were derived from a corpus of child speech which consists of recordings of Belgian Dutch children imitating Dutch (non-)words that were aurally presented to them (Verhoeven et al., 2016). The children were 6-year-olds differing in hearing status: one group consisted of children with normal hearing, while the other group consisted of children with hearing impairment. The latter contained children with a conventional hearing aid as well as children with a cochlear implant.

Stimuli and data selection

The database contains recordings of 36 monosyllabic (non-)words which consisted of a vowel nucleus with one of the 12 monophthongs of Belgian Standard Dutch, i.e. /i, ʏ, ɪ, ɛ, ɑ, ɔ, u, yː, eː, ø, aː, oː/ (Verhoeven, 2005). Specifically, 16 stimuli were existing Dutch words. The 20 other stimuli were non-words which respect the phonotactic structure of Dutch words. Each vowel occurred in three consonantal contexts: (i) /p_t/, (ii) /l_t/ and (iii) /t_r/. The overall structure of the non-words was compatible with the requirements of the Dutch phonological system.

These (non-)words had first been read aloud by a trained female phonetician and native speaker of Standard Belgian Dutch. The pitch contours of those vocalic productions are illustrated in Figure 1. It is clear that all the example words were realized phonetically with a fall and that there is no indication of list intonation in the examples. The recordings of these stimuli were played to the participating children who were asked to repeat the stimuli one by one. As such, the task is supposed to invoke new information status. All the recordings were made by means of a TASCAM DAT recorder and a head-mounted MicroMic II in a quiet room. The audio files were formatted to WAV-files by means of a TASCAM US 428 Digital Control Surface. The recording sessions with the children yielded a total of 7,985 children's imitations.

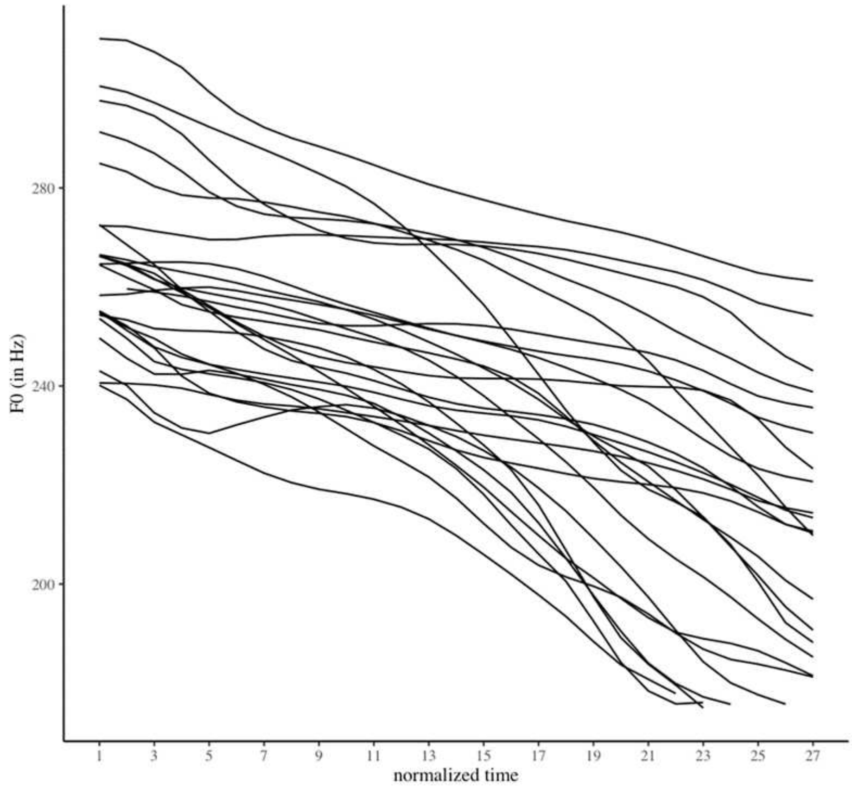


Figure 1: F0 contours in Hz of the stimuli presented aurally to the children.

In the first instance, the children's imitations were perceptually assessed by 6 trained phoneticians who identified the files in which the vowels were correct imitations of the target vowels. The perceptual assessment consisted in a multiple force choice task that was implemented in PRAAT (Boersma & Weenink, 2022). The listeners were aware of the hearing status (CI, HA or NH) and regional background of the children. They worked independently from each other. The judges were instructed to focus specifically on the vowel quality. They had to label the correctness of the vowel imitations as "correct", "incorOzbič & Kogovšekrect", or as "impossible to judge". Inter-rater agreement amounted to 97%. 7,261 vowels from 105 children were labelled as correct, i.e. 90.93% of the files. From these files, only those children were chosen who produced at least 2 repetitions of the corner vowels /i, aː, u/ and at least one repetition of all the vowels in any context. This amounted to a total of 5,557 files produced by 57 children. This selection was further reduced by only selecting the words with vowels following the voiceless stops in order to ensure that the F0 dynamics should be similar for every vowel in the corpus. The pitch tracking procedure did not provide any values for 9 vowels, probably due to the shortness of the window length, and these were also excluded from the analysis. The final selection consisted of 3,695 (non-)words.

Participants

The participants reflected in the corpus were 57 Belgian Dutch children born from native speakers of Belgian Standard Dutch. One group of 47 children had normal hearing on the basis of informal reports from parents and teachers. The mean chronological age of these children was 6 years, with a minimum of 5 and a maximum of 7 years. They all attended their first year of primary school and they had always lived in the region where they were born up to the point of data collection.

Besides children with normal hearing, there was a group of 5 children with a cochlear implant and 5 children with a conventional hearing aid. The hearing characteristics of the two groups of children with hearing impairment are summarised in Table 2 and Table 3:

Table 2: Characteristics of the selected CI children [CI: cochlear implant; HA: hearing aid; HL: hearing loss].

ID	Un-aided HL (in db)	Age at HA (in months)	HL with HA (in db)	Age at CI (in months)	Age at CI fitting (in months)	HL with CI (in db)	Device experience (in months)
CI1	120	1	120	7	8	27	66
CI2	115	2	113	10	12	25	62
CI3	93	5	47	17	18	35	56
CI4	117	4	107	5	6	17	67
CI5	112	2	58	19	21	30	53

Table 3: Characteristics of the selected HA children (HA: hearing aid; HL: hearing loss).

ID	Un-aided HL (in db)	Age at HA (in months)	HL with HA (in db)	Device experience (in months)
HA1	73	26	40	46
HA2	72	10	40	62
HA3	70	9	35	63
HA4	68	32	35	40
HA5	40	9	25	63

Acoustic analysis

The F0 of all the vowels was measured by means of a Python script. The calculation of F0 was done using the Parselmouth API (Jadoul et al., 2018) of PRAAT (Boersma & Weenink, 2022) via its standard auto-correlation algorithm. The maximum number of candidates was set to 15, the silence threshold to 0.03, the voicing threshold to 0.45, the octave cost to 0.01, the octave-jump cost to 0.35, the voiced/unvoiced cost to 0.14. In order to minimise the number of potential octave jumps, the pitch floor and ceiling were set at 175 Hz and 425 Hz respectively after visual inspection of the pitch contours with PRAAT's standard parameters. PRAAT's "Kill octave jumps" function was also applied. The F0 measurements were carried out in Hz, but a Lobanov-normalization procedure (Lobanov, 1971) was used to reduce potential effects of anatomical differences between children. However, the F0 contours in this paper are presented graphically on the Hz scale for ease of inspection.

All the vowels had been segmented by hand. Each vowel was then chunked into 10 frames, each frame representing one tenth of the total duration of the vowel. In addition, there was a frame overlap equal to two-thirds of a frame's duration, which means that each vowel was in effect subdivided into 27 frames. This resulted in 27 F0 measurements which made it possible to draw pitch contours and analyse their dynamics. The F0 value for each frame was taken as the mean of all the F0 measurements in that frame.

Statistical analysis

This study is interested in differences in the mean F0 associated with prominent syllable nuclei. In the first part of the analysis, the mean F0 over the entire vowel between the three groups of children was explored. Given the intrinsically hierarchical nature of these data, multi-level modelling was used as statistical technique.

Models of increasing complexity were built step-by-step by including random and fixed effects one after the other. After the inclusion of each of them, a likelihood ratio test was performed to establish whether the inclusion of that effect significantly improves the fit between the predicted and observed values. A model was preferred over another if the significance level of the model comparison is $p<0.05$. In the case of equally good fit with the data, the most parsimonious model was preferred. This method was based on Bates et al. (2015) and Baayen (2008). More details about the model comparison can be found in Appendix 2. The statistical analysis was carried out in R (R Core Team, 2021) and the R package *lme4* (Bates et al., 2015). *lmerTest* (Kuznetsova et al., 2015) was used to obtain p-values.

The final model is presented in Appendix 1. It included Consonant1 (C1), vowel identity, vowel duration as well as the interaction between Consonant1 and vowel and that between vowel and vowel duration. The inclusion of the children's hearing status did not significantly improve the model and was, therefore, excluded. The dependent variable was the mean F0 calculated over all frames of a vowel. As random effects, the basic model included random intercepts per child and per item nested within child.

In the second part of the analysis, this study aimed to investigate the dynamic F0 patterns within the vowels: this was achieved by means of Generalised Additive Mixed Modelling (henceforth: GAMM). This modelling technique allows more flexibility in dealing with non-linear patterns (Wieling, 2018) than polynomial regressions for instance. GAMMs are particularly well suited because they can model variation in height (e.g. vowels produced with different pitch heights) and in shape of trajectory (e.g. vowels produced with different F0 contours). Parametric terms help capture variation in height, and the smooth terms model the variation in shape. The statistical analysis was carried out in R (R Core Team, 2021) by means of the *mgcv* package (Wood, 2015). The *start_event* function of the *itsadug* package (van Rij et al., 2015) was used to order the dataset, and an autoregressive parameter of lag 1 was included which was obtained with a model without the autoregressive parameter, i.e. a ρ value of 0.945. The frame variable was centred so that the parametric terms are representative of the vowel at its centre. The outcome variable of the model are the F0 values normalised through a Lobanov-transformation (Lobanov, 1971). The model comparison consisted of χ^2 tests between a baseline and a more complex model via the *compareML* function of the *itsadug* package (van Rij et al., 2015). The model selection used is similar to the one described in Kirkham et al. (2019). Details can be found in Appendix 4.

The models involved in the model comparison were fitted with the maximum likelihood estimation. The visualization of the residual distribution showed two tails, which are indicative of a t-distribution. Consequently, the final model was fitted with a scaled-t distribution. This model was fitted with the fast restricted maximum likelihood estimation.

The final GAMM is presented in Appendix 3. It includes a single smooth of time and of vowel duration in order to model the main effect of time and duration on the F0 values. A tensor product interaction was included to capture the nonlinear interaction between the duration of the vocalic segment and time. A single smooth for the effect of child is included. A random reference smooth has been included to model the potential non-linear difference over time according to the time pattern of each individual children. By-children linear random slopes for the different prevocalic consonants (C1) have been included in order to allow the effect of C1 to vary between children. A random difference smooth was added, that is, a curve fit to the non-linear difference over time according to the duration of vowel with a by-effect of (ordered) hearing status and another one with a by-effect of C1. The duration of the vowel is included as a factor in the analysis to capture its effect.

Results

In the present study, a total number of 3,695 vowel productions were analysed acoustically for F0. In terms of average F0, the results of the multi-level modelling procedure in Table 4 show that there are F0 differences between the three speaker groups: the HA children have a somewhat higher F0 (284 Hz; SD=45) than the CI children (277 Hz; SD=38) children and children with normal hearing (270 Hz; SD=34). However, these differences are not significant.

Nevertheless, the statistical analysis revealed a significant fixed effect of both the identity of the vowels and of the identity of the prevocalic consonant (C1). This means that the average F0 varies significantly according to the C1 and the vowel (see Table 4). In terms of the effect of the prevocalic consonant on vowel F0, it is clear that the mean vowel F0 after /p/ is higher than after /t/ (277 Hz vs. 266 Hz respectively). As far as the mean F0 on the different vowels is concerned, the F0 differences are such that F0 is higher on the high vowels as compared to the low vowels: the vowel /aː/ has a mean F0 of 259 Hz, while /i/ and /u/ have higher mean F0 values, i.e. 290 Hz and 284 Hz respectively.[1]

Table 4: Results of the fixed effects of the best-fitting model [C1/p/ & vowel /ɑ/ = reference category]. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1. The results of the interaction effects are not reproduced here.

	Estimate	SE	df	*t*-value	*p*-value	
(Intercept)	0.27773	0.15959	2530.47648	1.74	0.08193	.
vowel duration	-1.42869	1.14618	2833.20236	-1.246	0.21269	
vowel /aː/	-0.87891	0.27992	2624.44372	-3.14	0.00171	**
vowel /ɛ/	-0.22077	0.22242	2606.15682	-0.993	0.321	
vowel /eː/	-0.27817	0.29811	2983.3732	-0.933	0.35084	
vowel /øː/	-0.57127	0.28486	2803.34279	-2.005	0.04501	*
vowel /ɪ/	0.12165	0.21401	2709.28656	0.568	0.5698	
vowel /i/	0.4443	0.20114	2511.09478	2.209	0.02727	*
vowel /ɔ/	0.06885	0.22285	2495.9368	0.309	0.75739	
vowel /u/	0.95945	0.20628	2586.82055	4.651	<0.001	***
vowel /oː/	-0.21534	0.28466	3075.27342	-0.756	0.44942	
vowel /ʏ/	0.26715	0.21378	2500.01094	1.25	0.21156	
vowel /yː/	0.6035	0.20624	2509.0167	2.926	0.00346	**
C1 /t/	-0.39664	0.09913	1436.90168	-4.001	<0.001	***

The second point of interest in this study has to do with the dynamics of the pitch contours in the three speaker groups. The non-normalised F0 contours in the three groups of children are illustrated in Figure 2.

[1] To obtain those values in Hz, the final model was refitted with the unnormalized F0 measurements in Hz.

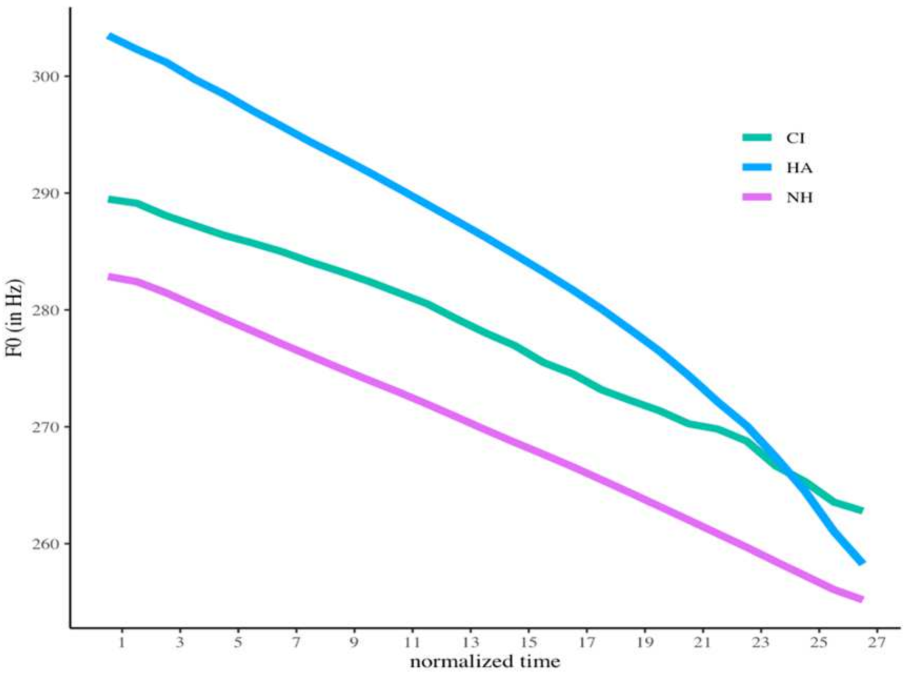


Figure 2: Average F0 contours in Hz for the three groups of children differing in hearing status [CI: cochlear implant; HA: hearing aid; NH: normally hearing].

In terms of the slope of the pitch contours in the speaker groups, there are clear slope differences. For the HA group is the slope of the fall 16.313 ST/sec (*SD*=10.133), while the slope for the CI group amounts to 11.196 ST/sec (*SD*=12.557) and 11.161 ST/sec (*SD*=8.319) for the NH group. Interestingly enough, the F0 slope of the HA group seems to consist of two phases. A first part situated over the first seven-nineth of the vowel has a slope of 13.427 ST/sec (*SD*=9.247) and is followed by an abruptly steeper decrease in F0 which amounts to 22.772 ST/sec (*SD*=20.879). On the contrary, the NH and CI groups display a more constant F0 slope throughout the vowel.

Looking at the variability in the F0 contours between the children, it can be seen in Figure 3 that there is a considerable amount of variability between children. The majority of the NH-children produce a prominence-lending accent realized as a fall similar to the fall produced in the stimuli which were presented to them. It should, however, be signalled that four children produce a rise late in the vowel. This might reasonably be attributed to a question intonation contour through which the child asks whether he or she is correctly performing the task. Visual inspection of the individual contours reveals that two children from the CI group (CI1 and CI5) produce a fall very much similar to the general pattern observed in the other groups and the original stimuli, while the three other children (CI2, CI3 and CI4) produce a rather flat F0 contour.

Note that the F0 contours displayed in Figures 1, 2 and 3 are time normalised. This means that the differences in vowel duration between each item are normalised. However, Figure 4 indicates that there is some variability in vowel durations which needs to be accounted for in the analysis. The CI group produces slightly shorter vowels (*M*= 163 ms; *SD*= 70) than the HA (*M*= 190 ms; *SD*= 85) and NH (*M*=183 ms; *SD*= 70) groups.

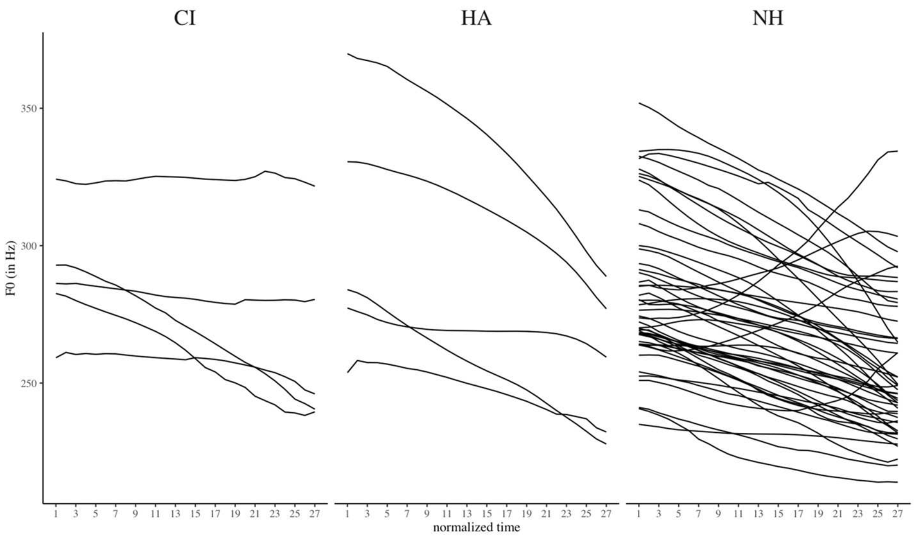


Figure 3: Average F0 contours in Hz per child and hearing status [CI: cochlear implant; HA: hearing aid; NH: normally hearing].

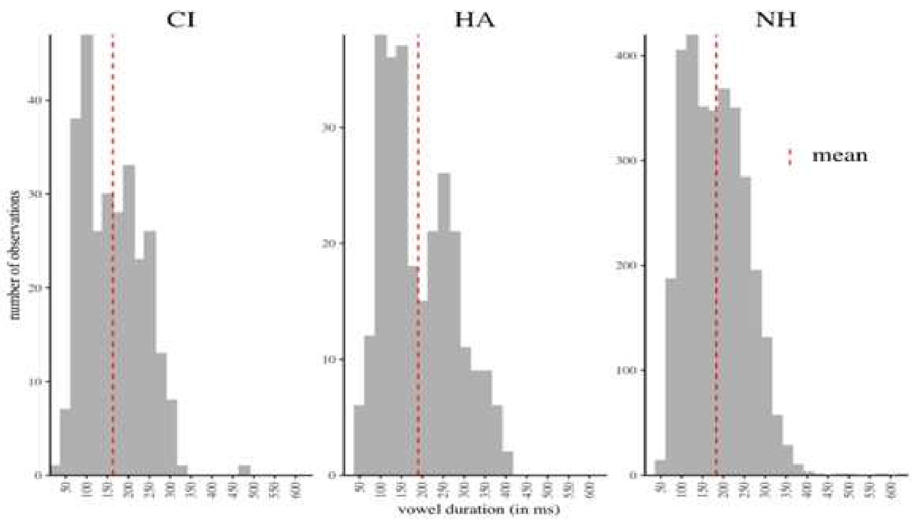


Figure 4: Distribution of vowel duration (in ms) per hearing status [CI: cochlear implant; HA: hearing aid; NH: normally hearing].

As far as the dynamics of the F0 contour are concerned, the results of the GAMM which fits the data best are reported in Table 5. There is a significant parametric effect of C1 (p<0.001).

As expected, the model shows a significant effect of time and of the duration of the vowel ($p<0.001$), their interaction is not significant after refitting the model with a *t*-distribution. Significant random smooths per subjects and random smooths over C1 are observed. Their effect is significant ($p<0.001$). A random reference smooth per subject over time is shown in the model and its effect is also significant ($p<0.001$). Tensor product interactions between time and duration of the vowel according to C1 are included in the model and their effect are significant for /t/ ($p<0.001$), but not for /p/.

As far as the research question of this paper is concerned, the effect of the interaction between duration of the vowel and time is significantly different between the CI and the HA children ($p<0.01$), but not between the CI and the NH. The non-significance of the latter can reasonably be attributed to the presence of two distinct patterns within the CI group. In fact, if the same model is refitted with the data of only the CI children who do not produce a clear fall, the group difference proves significant ($p<0.001$). It indicates that CI children might differ in the way they modulate F0 to mark prominence, but some of them might produce typical NH-like patterns. That individual pattern is levelled out in the analysis if all CI children are considered as a single group. In other words, the results cannot be generalized to all children, but they show that the modulation of F0 can be affected by CI use.

Generally speaking, it means that the effect of the non-linear interaction between time and duration of the vowel is different according to the hearing status of the child. It means that the effect of time modulated by the duration of the vowel has a different non-linear effect on the F0 values according to the hearing status of the child. GAMMs coefficients are per se difficult to interpret. The visualization of the values predicted by the model are therefore presented in Figure 5. We observe in Figure 5 that the CI group has a relatively flatter F0 contour than the two other groups. On the contrary, the HA and NH groups produce a clearly marked fall. During the first part of the vowel, the HA and NH contours are, after normalization, quite close to each other. However, towards the end of the vowel, the F0 contour becomes more abrupt in the HA group than in the NH group.

Discussion

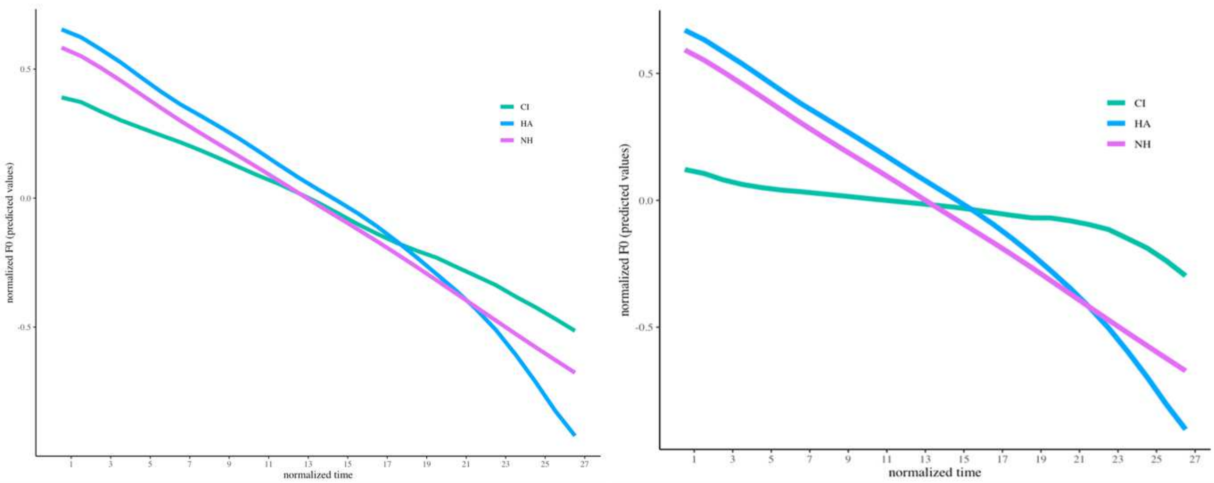
The key research question of this study was whether children differing in hearing status differ in the (dynamics of) F0 associated with prominence after voiceless plosives within the nucleus of monosyllabic words. This study was carried out on the speech of 6-year-old hearing-impaired children who had been fitted with either a cochlear implant (N=5) or a conventional hearing aid (N=5) in comparison with a control group which consisted of 47 children with normal hearing. The pitch values were extracted via PRAAT inside time-normalised overlapping frames. The mean F0 was analysed by means of multilevel modelling, while the overall contours for the three groups were modelled via GAMMs. From the results, it appears that there are differences in the mean F0 and the dynamics of the F0 contours between the three groups. However, not all the observed differences are significant.

As far as the mean F0 is concerned, visual inspection of the data reveals that the mean F0 associated with prominence in children with hearing impairment is somewhat higher

Table 5: Summary of the GAMM created to compare the F0 contours of the CI, HA and NH groups. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1. [C1/p/ = reference category].

A. Parametric coefficients	Estimate	SE	*t*-value	*p*-value
Intercept	0.0843	0.0396	2.129	0.0333*
C1/t/	-0.30483	0.05428	-5.615	<0.001***
B. Smooth terms	edf	Ref.df	F-value	*p*-value
s(total_duration)	4.008	5.006	60.425	<0.001***
s(time_centred)	3.711	4.204	114.866	<0.001***
ti(Time_centred, total_duration)	2.05	2.905	0.242	0.86855
s(name, bs="re")	0.001	56	0	0.16922
s(name, C1, bs="re"):C1/p/	117.9	512	0.993	<0.001***
s(name, C1, bs="re"):C1/t/	289.2	512	1.886	<0.001***
s(Time_centred, name, bs="fs",m=1)	326.3	512	2.616	<0.001***
ti(Time_centred, total_duration):HA	2.599	3.279	3.656	0.00961**
ti(Time_centred, total_duration):NH	7.12	8.897	1.4	0.18368
ti(Time_centred, total_duration):C1/p/	2.286	2.672	1.86	0.12082
ti(Time_centred, total_duration):C1/t/	5.177	6.242	5.626	<0.001***

Rank: 1695/1696
R-sq. (adj) = 0.387
Deviance explained = 36.1%
fREML= 48695
Scale est. = 1
n = 99028



(a) *(b)*

Figure 5: Time course of F0 (in Hz) per hearing status (predicted values) [CI: cochlear implant; HA: hearing aid; NH: normally hearing]. (a) Full corpus; (b) CI1 and CI5 excluded.

than in children with normal hearing. This is consistent with previous research on hearing impairment which indicates that higher F0 values are characteristic of the speech of both hearing-impaired adults (Mora et al., 2010) and children (Angelocci et al., 1964; Higgins et al., 2003; Dehqan and Scherer, 2011). It should however be pointed out that the differences found in the present study are statistically not significant, so these results do not provide strong confirmation of the findings of previous research. It is not clear how to account for the non-significance of the findings in this paper, but this may be due to a combination of very rigorous statistical testing (as compared to other research), the small number of participants in some of the subgroups and the high variability in these groups.

In addition, the statistical analysis revealed a significant effect of both the vowel and the prevocalic consonant on the mean F0. In terms of vowel identity as a factor it was clear that the high vowels have a significantly higher F0 than the low vowels. The difference amounts on average to 23 Hz. This finding is consistent with many studies which have documented such F0 differences between high and low vowels in a large number of languages and in different groups of speakers (among others: Ladefoged 1964; Traunmüller, 1981; Whalen & Levitt, 1994; Van Hoof & Verhoeven, 2011). However, this is the first time that a robust effect of instrinsic vowel F0 has been documented in such a large group of child speakers, both with and without hearing impairment. This may open interesting perspectives for further research, particularly into the question of whether intrinsic F0 can be actively controlled by speakers for communicative purposes such as to perceptually enhance vowel contrasts. However, it should be noted that other sources of variability that the present carefully controlled task is unable to account for might be present in more ecological contexts.

Furthermore, the analysis also showed a significant effect of the prevocalic consonant on the F0 of the vowel nucleus. /p/ as a C1 triggers higher F0 values in the vowel than /t/ i.e. 277 Hz vs. 266 Hz. Although the differential effect of plosive voicing on the F0 of adjacent vowels is well-known and has been documented before (House & Fairbanks, 1953; Lehiste & Peterson, 1961; Xu & Xu, 2021), the finding of a significant difference in place of articulation of the voiceless plosives is new: the data in House & Fairbanks (1953) and Lehiste & Peterson (1961) show F0 differences of around 1 Hz between the two places of articulation.

As far as the pitch configuration is concerned, the modelling procedure indicates that the general pitch configuration associated with prominence in this study is a prominence-lending F0 fall which is the phonetic manifestation of an underlying rise-fall nuclear accent contour. Visual inspection of the F0 contours in the different groups of children in Figure 3 reveals clear differences between the three groups which are found to be significant at a statistical level.

From Figure 5, it is clear that the slope of the falls associated with prominence in CI children is significantly shallower than in the other groups. This indicates that there is less modulation of F0 in the CI group than in the HA and NH groups. The smaller modulation of F0 by CI users is consistent with previous research on the speech production of CI users (e.g. Hide, 2014, De Clerck et al., 2018) which describes the speech of CI speakers as more monotonous and this might indicate that the lack of detail in the auditory input provided to children fitted with CI is mirrored by a lack of F0 modulation produced by them (cfr. Moore, 2003; Green et al., 2004; Fu & Nogaki, 2005; O'Halpin, 2010). It should, however, be emphasised that there are differences between individual CI children in this respect. Although most of the CI children in this study produce a relatively flat F0 contour, there are some CI children who produce a fall which is very similar to their NH peers. The reasons for these individual differences are not immediately clear since there are no specific auditory characteristics which can explain these differences in F0 modulation. It can nevertheless be seen from this study that the acquisition of F0 contours similar to that of children with normal hearing is not impossible to achieve by children fitted with a CI. The CI's greater variability in speech production in general is also consistent with the present findings.

When it comes to the HA group, a more similar F0 contour to the NH group is found: in all the participants with HA the F0 drops without exception towards the end of the vowel nucleus at a rate which is very similar to the children with normal hearing. However, towards the end of the vowel, the F0 tends to drop more steeply. The cause of this sudden drop in F0 is not clear but might be attributed to a lower degree of control over phonation towards the end of the vowel. The finely-tuned control of F0 needed for the realisation of prominence might be achieved by HA children throughout the first part of the vowel but might be difficult to maintain until the end. Note also in Figure 4 that the HA group produces, on average, slightly longer vowels than the two other groups.

This difficulty to maintain fine control over phonation throughout the vowel might lead to a change in voice quality, such as creaky voice, towards the end of the vowel and this might be the cause of the observed lower F0 values towards the end of the vowels produced by the HA children. This would be in accordance with what Van Lierde et al. (2005), Valero Garcia et al. (2010) and Baudonck et al. (2011) suggest about the voice quality of hearing-impaired speakers, especially for HA users. In fact, Valero Garcia et al. (2010) suggest that the use of analogue hearing aids, contrary to digital hearing aids, is associated with worse voice quality than that of cochlear implanted children. Similarly, Van Lierde et al. (2005) find that the two groups differ with respect to the norm in terms of jitter, whereby the CI group exhibits lower jitter and the HA group higher jitter. Based on a GRBAS perceptual rating of voice quality Baudonck et al. (2011) found that HA and CI (both unilaterally and bilaterally implanted) exhibit a strained and unstable phonation. They also point out that the HA obtained the worst scores on the GRBAS scale.

The above-mentioned differences between groups are observable as an effect of the hearing status of the children on the non-linear interaction between time and duration. It shows that the effect of time on the normalised pitch values is different between groups, but the effect of time is modulated by the duration of the vowels. As such, the differences in vowel duration between groups are accounted for in the analysis. It means that the different groups do not differ in height of F0 but in the non-linear F0 pattern, i.e. the shape of the contour.

A difference in F0 contour can also be observed between the two hearing-impaired groups. It can reasonably be attributed to device use. Hearing aids produce a more faithful input of F0, which is not the case for CI. The potential sound distortion caused by the amplification of given frequencies by a conventional hearing aid (Metcalfe, 2017) might distort the original input to a certain extent, but the resolution would still be higher than that provided by the limited number of electrodes of a CI (Moore, 2003; Green et al., 2004; Fu & Nogaki, 2005; O'Halpin, 2010). This might consequently make the acquisition of an NH-like pattern easier for HA than for CI children. The data nevertheless show that, despite the low resolution of the signal provided by CI, some children manage to acquire an NH-like pattern by the age of 6. Over time, CI children might learn to infer information about F0 from the higher harmonics because the degraded input might make it difficult to rely mainly on F0 itself. That might be an explanation for the delayed, but not impossible, acquisition of NH-like contours by CI children.

In a nutshell, the three hearing groups produce the F0 modulations needed for the realisation of prominence differently. The children fitted with a cochlear implant produce a less steep fall. This is in agreement with research describing CI users' speech as monotonous as a result of the lack of detail in acoustic input provided by the implant. On the contrary, the HA group produce more naturally modulated F0 patterns which are close to the patterns of their NH peers apart from a sharp decrease in F0 towards the end of the vowels.

Limitations of this study

This research sheds light on the modulation of F0 by three groups of children differing in hearing status, but it may be useful to point out some limitations arising from the methodological choices that were made.

Firstly, the nature of a word and non-word repetition task is anything but ecological. In fact, the design of the experiment aimed to elicit speech in laboratory conditions in order to control for experimental variables. Secondly, to enhance within-children data homogeneity and reduce computational cost, a further selection was applied on the original corpus (Verhoeven et al., 2016). As a result, the data selection significantly reduces the sample size and caution must be taken when interpreting the results of this much more restricted dataset. The structure of the data is also somewhat unbalanced. Thirdly, the parametric settings of the F0 tracking algorithm were set in order to reasonably include the pitch range of all the participating children. It does not entail automatically that those settings are ideal for each child and that the F0 tracking does not make mistakes. Furthermore, this paper focuses on the description of detailed modulations in the speech of normally hearing and hearing-impaired children but does not engage with the larger F0 modulations to be expected in longer and more complex utterances. The extent to which these detailed F0 modulations participate in the modulations of a larger prosodic structure is unclear.

Another limitation is that the speech materials are made up of repetitions of monosyllables and as such word-stress coincides with sentence stress in utterance-final position so that effects of end-of-utterance on F0 can be expected such as the sudden drop in F0 in HA children.

It should also be noted that the GAMM analysis does not engage with vowel effects for computational cost reasons. We nonetheless believe that a potential effect of vowel category should be reflected in a height effect as demonstrated by the multi-level modelling analysis, but it is not expected to have the shape of a trajectory effect. Moreover, the functional relevance of the differences in F0 dynamics described here above are still unclear. In other words, it is not clear whether a less steep nuclear fall hampers the intelligibility of CI users. Given the individual differences within a single group, it would be of interest to collect more data. Longitudinal data might also help understand whether the production of an NH-like pattern by CI users depends on individual learning trajectories.

Conclusions

This paper investigated the dynamics of F0 contours in prominent vowel nuclei after voiceless plosives in two groups of hearing-impaired children and one group of normally hearing children. The modelling procedure indicated clear differences in F0 both in terms of its overall mean in the vowel nucleus as well as in the shape of the contours. The overall mean F0 in the vowel nucleus was somewhat higher in children with hearing impairment than in children with normal hearing. Although this is consistent with previous findings on hearing impaired speech, the difference in this study was not significant. Additionally, this study revealed that most of the NH children realise the prominent syllable in this specific task through a pointed-hat pattern realised phonetically as a fall. On the one hand, HA children produce F0 contours similar to those produced by NH children, even if HA children tend to produce a steeper fall towards the end of the vowel. This might be indicative of a difficulty to maintain fine-tuned control over the total vowel duration. On the other hand, CI children often produce a rather flat F0 contour. This might be a consequence of the lack of spectral details provided by the CI device. However, some CI children show F0 contours similar to those of their NH peers.

Acknowledgements

This work was supported by the Research Foundation in Flanders (FWO) [grant G004321N].

References

Angelocci, A. A., Kopp, G. A., & Holbrook, A. (1964). The Vowel Formants of Deaf and Normal-Hearing Eleven- to Fourteen-Year-Old Boys. *Journal of Speech and Hearing Disorders, 29*(2), 156-170. https://doi.org/10.1044/jshd.2902.156

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge ; New York: Cambridge University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. http://dx.doi.org/10.18637/jss.v067.i01

Baudonck, N., Dhooge, I., D'haeseleer, E., & Van Lierde, K. (2010). A comparison of the consonant production between Dutch children using cochlear implants and children using hearing aids. *International Journal of Pediatric Otorhinolaryngology, 74*(4), 416-421. http://dx.doi.org/10.1016/j.ijporl.2010.01.017

Baudonck, N., Lierde, K. V., D'haeseleer, E., & Dhooge, I. (2011). A comparison of the perceptual evaluation of speech production between bilaterally implanted children, unilaterally implanted children, children using hearing aids, and normal-hearing children. *International Journal of Audiology, 50*(12), 912-919. https://doi.org/10.3109/14992027.2011.605803

Boersma, P. & Weenink, D. (2022). *Praat: doing phonetics by computer* (version 6.2.06).

De Clerck, I., Pettinato, M., Gillis, S., Verhoeven, J., & Gillis, S. (2018). Prosodic modulation in the babble of cochlear implanted and normally hearing infants: A perceptual study using a visual analogue scale. *First Language, 38*(5), 481-502. http://dx.doi.org/10.1177/0142723718773957

Dehqan, A., & Scherer, R. C. (2011). Objective Voice Analysis of Boys with Profound Hearing Loss. *Journal of Voice, 25*(2), e61-e65. https://doi.org/10.1016/j.jvoice.2010.08.006

Dettman, S. J., Dowell, R. C., Choo, D., Arnott, W., Abrahams, Y., Davis, A., Dornan, D., Leigh, J., Constantinescu, G., & Cowan, R. (2016). Long-term communication outcomes for children receiving cochlear implants younger than 12 months: A multicenter study. *Otology & Neurotology, 37*(2), e82-e95. http://dx.doi.org/10.1097/MAO.0000000000000915

Fu, Q.-J. & Nogaki, G. (2005). Noise Susceptibility of Cochlear Implant Users: The Role of Spectral Resolution and Smearing. *Journal of the Association for Research in Otolaryngology, 6*(1), 19-27. http://dx.doi.org/10.1007/s10162-004-5024-3

Geffner, D. (1980). Feature characteristics of spontaneous speech production in young deaf children. *Journal of Communication Disorder, 13*(6), 443-454.

Gold, T. (1980). Speech production in hearing-impaired children. *Journal of Communication Disorders, 13*(6), 397-418. https://doi.org/10.1016/0021-9924(80)90042-8

Green, T., Faulkner, A., & Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *The Journal of the Acoustical Society of America, 116*(4), 2298-2310. http://dx.doi.org/10.1121/1.1785611

't Hart, J. & Collier, R.P.G. (1979). On the interaction of accentuation and intonation in Dutch. In Fischer-Jørgensen, E., Rischel, J., & Thorsen, N. (Eds.), *Proceedings of the Ninth International Congress of Phonetic Sciences,* Held in Copenhagen, 6-11 August 1979 (pp. 395–402). University of Copenhagen.

´t Hart, J., Collier, R., & Cohen, A. (1990). *A Perceptual Study of Intonation: An Experimental Phonetic Approach to Speech Melody.* Cambridge ; New York: Cambridge University Press.

Hide, O. (2014). *Acoustic Features of Speech by Young Cochlear Implant Users: A Comparison with Normal-Hearing and Hearing-Aided Age Mates* [PhD thesis, University of Antwerp].

Higgins, M. B., McCleary, E. A., Carney, A. E., & Schulte, L. (2003). Longitudinal changes in children's speech and voice physiology after cochlear implantation. *Ear and Hearing, 24*(1), 48-70. https://doi.org/10.1097/01.AUD.0000051846.71105.AF

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25*, 105-113.

Hudgins, C. V., & Numbers, F. C. (1942). An investigation of the intelligibility of the speech of the deaf. *Genetic Psychology Monographs, 25*, 289-392.

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics, 71*, 1-15. https://doi.org/10.1016/j.wocn.2018.07.001

Kirkham, S., Nance, C., Littlewood, B., Lightfoot, K., & Groarke, E. (2019). Dialect variation in formant dynamics: The acoustics of lateral and vowel sequences in Manchester and Liverpool English. *The Journal of the Acoustical Society of America, 145*(2), 784-794. https://doi.org/10.1121/1.5089886

Korver, A. M. H., Smith, R. J. H., Van Camp, G., Schleiss, M. R., Bitner-Glindzicz, M. A. K., Lustig, L. R., Usami, S.-I., & Boudewyns, A. N. (2017). Congenital hearing loss. *Nature Reviews. Disease Primers, 3*:16094. http://dx.doi.org/10.1038/nrdp.2016.94

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2015). *Package 'lmertest'*. R package version 2, 734.

Ladefoged, P. (1964). *A phonetic study of West African languages.* Cambridge: Cambridge University Press.

Lehiste, I., & Peterson, G. E. (1961). Some Basic Considerations in the Analysis of Intonation. *The Journal of the Acoustical Society of America, 33*, 419-425. https://doi.org/10.1121/1.1908681

Lenden, J. M. & Flipsen, P. (2007). Prosody and voice characteristics of children with cochlear implants. *Journal of Communication Disorders, 40*(1), 66-81. http://dx.doi.org/10.1016/j.jcomdis.2006.04.004

Levine, D., Strother-Garcia, K., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Language development in the first year of life: What deaf children might be missing before cochlear implantation. *Otology & Neurotology, 37*(2), e56-e62.

Lobanov, B. M. (1971). Classification of Russian Vowels Spoken by Different Speakers. *The Journal of the Acoustical Society of America, 49*(2B), 606-608. http://dx.doi.org/10/cqm563

Maassen, B. (1986). Marking word boundaries to improve the intelligibility of the speech of the deaf. *Journal of Speech, Language, and Hearing Research, 29*(2), 227-230.

Markides, A. (1970). The speech of deaf and partially-hearing children with special reference to factors affecting intelligibility. *British Journal of Disorders of Communication, 5*(2), 126-140.

Metcalfe, T. (2017). *Perception of Speech, Music and Emotion by Hearing-Impaired Listeners* [PhD thesis, University of Sheffield].

Moore, B. C. (2003). Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otology & neurotology, 24*(2), 243-254. http://dx.doi.org/10.1097/00129492-200303000-00019

Mora, R., Jankowska, B., Guastini, L., Santomauro, V., Dellepiane, M., Crippa, B., & Salami, A. (2010). Computerized Voice Therapy in Hypofunctional Dysphonia. *Journal of Otolaryngology - Head & Neck Surgery, 39*, 615-621. https://doi.org/10.2310/7070.2010.090209

Nickerson, R. S. (1975). Characteristics of the Speech of Deaf Persons. *Volta Review, 77*(6), 342-62.

Nober, E. H. (1967). Articulation of the Deaf. *Exceptional Children, 33*(9), 611–621. https://doi.org/10.1177/001440296703300902

O'Halpin, R. (1997). Contrastive Stress in the Speech of Profoundly Deaf Children: A Preliminary Investigation. *Volta Review, 99*(2), 89-105.

O'Halpin, R. (2010). *The Perception and Production of Stress and Intonation by Children with Cochlear Implants* [PhD thesis, University of London].

Osberger, M. J., & Mcgarr, N. S. (1982). Speech Production Characteristics of the Hearing Impaired. In N. J. Lass (Ed.), *Speech and Language* (Vol. 8, pp. 221–283). https://doi.org/10.1016/B978-0-12-608608-9.50013-9.

Ozbič, M., & Kogovšek, D. (2008). An acoustic comparison of formant frequencies in individuals with normal hearing, profound and severe hearing impairment. *Investigationes Linguisticae, 16*, 150–162. https://doi.org/10.14746/il.2008.16.13

Ozbič, M., & Kogovšek, D. (2010). Vowel formant values in hearing and hearing-impaired children: A discriminant analysis. *Deafness & Education International, 12*(2), 99-128. https://doi.org/10.1179/146431510X12626982043804

R Core Team, (2021). *R: A Language and Environment for Statistical Computing.* Vienna, Austria. Retrieved from https://www.R-project.org/

Smith, C. R. (1975). Residual hearing and speech production in deaf children. *Journal of speech and hearing research, 18*(4), 795-811.

Stein, D. M. (1980). *A study of articulatory characteristics of deaf talkers* [PhD thesis, The University of Iowa].

Stevens, K. N., Nickerson, R. S., Boothroyd, A., & Rollins, A. M. (1976). Assessment of Nasalization in the Speech of Deaf Children. *Journal of Speech and Hearing Research, 19*(2), 393–416. https://doi.org/10.1044/jshr.1902.393

Sussman, H., & Hernandez, M. (1979). A spectrographic analysis of the suprasegmental aspects of the speech of hearing-impaired adolescents. *Audiology and Hearing Education, 5*, 12-16.

Terken, J., & Hermes, D. (2000). The Perception of Prosodic Prominence. In M. Horne (Ed.) *Prosody: Theory and Experiment, studies presented to Gösta Bruce* (pp. 89–127). Dordrecht ; Boston ; London: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-9413-4_5

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *The Journal of the Acoustical Society of America, 69*(5), 1465-1475. https://doi.org/10/c2vbc9

Valero Garcia, J., Rovira, J. M. V., & Sanvicens, L. G. (2010). The influence of the auditory prosthesis type on deaf children's voice quality. *International Journal of Pediatric Otorhinolaryngology, 74*(8), 843-848. http://dx.doi.org/10.1016/j.ijporl.2010.04.011

van Geel, R. C. (1983). *Pitch inflection in electrolaryngeal speech.* [PhD thesis, University Utrecht].

Van Hoof, S., & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *Journal of Phonetics, 39*(2), 168-177. https://doi.org/10/bnth37

Van Lierde, K. M., Vinck, B. M., Baudonck, N., De Vel, E., & Dhooge, I. (2005). Comparison of the overall intelligibility, articulation, resonance, and voice characteristics between children using cochlear implants and those using bilateral hearing aids: A pilot study. *International Journal of Audiology, 44*(8), 452-465. http://dx.doi.org/10.1080/14992020500189146

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, D. (2015). *Itsadug: Interpreting time series and autocorrelated data using GAMMs.*

Vanormelingen, L., De Maeyer, S., & Gillis, S. (2016). A comparison of maternal and child language in normally-hearing and hearing-impaired children with cochlear implants. *Language, Interaction and Acquisition, 7*(2), 145-179. http://dx.doi.org/10.1075/lia.7.2.01van

Verhoeven, J. (2005). Belgian Standard Dutch. *Journal of the International Phonetic Association, 35*(2), 243-247. http://dx.doi.org/10.1017/S0025100305002173

Verhoeven, J., Hide, O., De Maeyer, S., Gillis, S., & Gillis, S. (2016). Hearing impairment and vowel production. A comparison between normally hearing, hearing-aided and cochlear implanted Dutch children. *Journal of Communication Disorders, 59*, 24-39. http://dx.doi.org/10/f787b7

Whalen, D., & Levitt, A. (1994). The universality of intrinsic F0 of vowels. *Haskins Laboratories Status Report on Speech Research*, SR-117/118: 1-14.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86-116. http://dx.doi.org/10.1016/j.wocn.2018.03.002

Wood, S. (2015). *Package 'mgcv'.* R package version, 1(29), 729.

Xu, Y., & Xu, A. (2021). Consonantal F0 perturbation in American English involves multiple mechanisms. *The Journal of the Acoustical Society of America, 149*(4), 2877-2895. https://doi.org/10.1121/10.0004239

Appendix

1. Listing - Final multi-level model

```
# Fixed effect of the duration of the vowel on the average
# normalized F0
final_model <- lmer(mean_z_pitch   (total_duration

# Fixed effect of prevocalic consonant (/p/ vs /t/)
+ C1)

# Fixed effect of the vowel category and interaction with C1 and
# duration of the vowel
* vowel

# Random effects of child and item nested within child
+(1|name/C1_V))
```

2. Model comparison - Multi-level modelling analysis

Table 6: Model comparison [*significantly improves data fit or more parsimonious model].

	Added term	Baseline model	Compared model	Npar	AIC	BIC	Loglik	Deviance	χ^2	df	*p*-value	Comparison output	Retained model *
Null Model with random intercept per child(M1)	Duration fixed effect (M2)	M1	M2	4	8190.4	8215.3	-4091.2	8182.4	753.72	1	<0.001	Random intercept per vowel needed	M2
	Duration fixed effect (M3)	M2	M3	5	8052.4	8083.5	-4021.2	8042.4	140	1	<0.001	Duration fixed effect needed	M3
	C1 fixed effect (M4)	M2	M4	6	7956.0	7993.3	-3972.0	7944.0	98.454	1	<0.001	C1 fixed effect needed	M4
	V fixed effect (M5)	M4	M5	17	7449.6	7555.2	-3707.8	7415.6	528.38	11	<0.001	V fixed effect needed	M5
	C1 * V interaction (M6)	M5	M6	28	7414.7	7588.7	-3679.3	7358.7	56.925	11	<0.001	C1*V interaction needed	M6
	C1 * Duration interaction (M7)	M6	M7	29	7410.3	7590.5	-3676.1	7352.3	6.394	1	0.01145	C1 * Duration interaction needed	M7
	Duration * vowel interaction (M8)	M7	M8	51	7429.0	7745.9	-3663.5	7327.0	25.296	22	0.2831	Duration * vowel interaction not needed	M7
	Hearing status fixed effect (M10)	M6	M10	41	7428.7	7683.5	-3673.4	7346.7	0.2745	2	0.8718	Equivalent data fit	M6

3. Listing - Final generalised additive mixed model

```
# Parametric term of C1 (i.e. fixed effect) on normalized pitch:
final_model <- bam(z_pitch   C1

# Single smooth (i.e. random intercept) for the effect of
# time_centred and total_duration
+s(total_duration)
+s(time_centred)

# Tensor product interaction of time_centred and total_duration
#(i.e. interaction between variables)
+ti(Time_centred, total_duration)

# Single smooth (i.e. random intercept) for the effect of child
+s(name, bs="re")

# Difference smooth (i.e. random slope) for the effect of child
# per C1
+s(name, C1, bs="re")

# Random smooth for the non-linear effect of time
# according to general pattern of each speaker
+s(Time_centred, name, bs="fs",m=1)

# Tensor product interaction for the effect of child per C1
# and hearing status
+ti(Time_centred, total_duration, by= C1)
+ti(Time_centred, total_duration, by= Hearing status))
```

Table 7: Model comparison [*significantly improves data fit or more parsimonious model].

	Added term		Removed term	Baseline model	Compared model	Score	Edf	Difference	Df	*p*.value	Comparison output	Retained model*
Null Model (M1)	Model single and random smooths per child (M1B)			M1	M1B	-66425.34	11	195341.267	3	<0.001	Single and random smooths per child needed	M1B
	With C1 predictor (M2A)			M1B	M2A	-70447.85	23	4022.510	12	<0.001	C1 predictor needed	M2A
		With smooth term only M2B)		M2A	M2B	-70447.85	23	14.500	1	<0.001	Parametric term needed	M2A
		With parametric term only (M2C)		M2A	M2C	-70447.85	23	3652.534	13	<0.001	Smooth terms needed	M2A
			Without tensor product interaction (M2A1)	M2A	M2A1	-70447.85	23	56.404	6	<0.001	Tensor product interaction needed	M2A
			Without single smooth (M2A2)	M2A	M2A2	-70449.01	22	-1.162	1	Only small difference in ML	Equivalent data fit	M2A2
			Without random smooth (M2A3)	M2A	M2A3	-70447.85	23	1923.328	4	<0.001	Random smooth needed	M2A
	With hearing status predictor (M3A)			M2A2	M3A	-70596.66	33	147.650	11	<0.001	Hearing status predictor needed	M3A
		With smooth term only (M3B)		M3A	M3B	-70572.56	35	0.458	2	0.633	Equivalent data fit	M3B
		With parametric term only (M3C)		M3A	M3C	-70572.56	35	124.263	11	<0.001	Smooth terms needed	M3A
				M3B	M3C	-70572.1	33	123.806	9	<0.001		M3B
			Without tensor product interaction (M3B1)	M3B	M3B1	-70572.10	33	122.932	6	<0.001	Tensor product interaction needed	M3B
			Without single smooth (M3B2)	M3B	M3B2	-70572.95	32	-0.850	1	Only small difference in ML...	Equivalent data fit	M3B2
			Without random smooth (M3B3)	M3B	M3B3	-70572.11	29	-0.011	4	Only small difference in ML...	Equivalent data fit	M3B3
			Without random and single smooths (M3B4)	M3B4	M3B2	-70572.95	32	0.973	4	0.746	Equivalent data fit	M3B4
				M3B4	M3B3	-70572.11	29	0.134	1	0.604		

STEM-, SPRAAK- EN TAALPATHOLOGIE

Vol. 27, 2022, Special Issue, pp. 93-134
Special Issue for Steven Gillis

Gepubliceerd online: 5 november 2022

University of Groningen Press
https://doi.org/10.21827/32.8310/2022-SG-93

The language of school writing: a developmental comparison of genres across the school years

Leah Haim & Dorit Ravid

School of Education, Tel Aviv University, Israel

Samenvatting

Schrijven is de voornaamste route voor het leren, verwerken, organiseren, opslaan en ophalen van informatie in de schooljaren. Het beheersen van de structuren en functies van teksten uit verschillende genres is een van de belangrijkste doelen van geletterdheid in het onderwijs. De teksten die leerlingen construeren, creëren een ideal domein voor het tonen van hun taalvaardigheden en voor het onderzoeken van de constructie van verschillende typen tekst, specifiek voor verschillende genres. Deze studie vergelijkt de taalkundige constructies die horen bij verklarende (*expository*) teksten (waar het gaat om argumentatie en overreding bij sociale en conceptuele kwesties) met informatieve teksten, die feitelijke (en minder controversiële) verschijnselen beschrijven. Het onderzoek biedt een perspectief op de schrijfvaardigheden van Hebreeuws sprekende leerlingen in het primair en voortgezet onderwijs, vergeleken met volwassenen. De deelnemers waren 547 leerlingen en volwassenen die vervolgonderwijs hadden gevolgd. Elke deelnemer schreef een informatieve tekst en een argumenterende tekst, wat leidde tot een totaal van 1094 teksten. Er werden drie soorten variabelen onderzocht: lexicale, morfo-syntactische en syntactische. We vonden dat complex lexicon en complexe syntaxis in de teksten toenamen tijdens de schooljaren, met specifieke structuren die karakteristiek zijn voor genres, al naar gelang de positie en de aard van de twee genres die we onderzochten. We vonden ook dat al deze componenten pas op volwassen leeftijd optimaal gebruikt werden, als een culminatie van de periode van latere taalontwikkeling. Onze resultaten impliceren dat informatieve en verklarende teksten inderdaad verschillend zijn in hun kenmerken, en dat het niet-deskundigen vele jaren van taalkundige en cognitieve ontwikkeling aan de ene kant en instructie op school en ervaring aan de andere kant kost om het niveau van kwalitatief academisch schrijven te bereiken.

Abstract

Writing is the highway to learning, processing, organizing, storing and retrieving information during the school years. Gaining command of the structures and functions of

Correspondentieadres:
Dorit Ravid
School of Education
Tel Aviv University
Ramat Aviv 69978, Tel Aviv
Israel
E-mail: doritr@tauex.tau.ac.il

texts of different genres is one of the main goals of linguistic literacy in education. The texts that school-goers construct provide optimal hunting grounds for unveiling their linguistic abilities during the genre-specific construction of different text types. The current study examines the linguistic constructions typical of expository texts, characterized by argumentation or persuasion regarding social / conceptual issues, versus informative texts, describing factual (or less controversial) phenomena. This examination constitutes a window on Hebrew-speaking students' developing writing abilities in elementary, middle and high school, compared with adults. Participants were 547 students and educated adults with post-high school education. Each participant wrote an informative text and an expository text, yielding a total of 1,094 texts. Three types of variables were examined: lexical, morpho-syntactic and syntactic. We found that complex lexicon and syntax in discourse increased in prevalence across the school years, with specific structures being genre-typical as befits the stance and character of the two genres under investigation. We also found that it was only in adulthood that all of these components were employed in optimal fashion, at the culmination of the period of Later Language Development. Our results imply that informative and expository texts are indeed distinct in their characteristics, and that it takes many years of internal linguistic and cognitive development, on the one hand, and schooling instruction and experience, on the other hand, to achieve qualitative academic writing in non-experts.

Introduction

Gaining command of the structures and functions of texts of different genres is one of the main goals of linguistic literacy in education. Expository and informative texts constitute the prevalent genres in academic reading and writing, underscoring the need to investigate students' developing expressive linguistic abilities in these genres (Beers & Nagy, 2011). Much psycholinguistic research on the development of writing skills has focused on the contrast between narrative and expository writing (Berman, 2009; Nippold & Scott, 2010). The current study turns to the linguistic characteristics of Hebrew informative versus expository writing as a window on language and literacy skills across the school years and beyond, during the period known as Later Language Development (Berman, 2005).

Later language development

Later language development, which takes place between the ages of 9 years to adulthood, ushers in mature native language proficiency (Berman, 2004, 2016; Berman & Ravid, 2008; Nippold, 2016). This is a time of great changes in the individual's brain structures and functions (Fuhrmann, Knoll, & Blakemore, 2015; Kadosh, Linden, & Lau, 2013; Paus, 2005), characterized by growing command of socio-cognitive and executive control abilities (Lecce et al., 2017; Osterhaus, Koerber, & Sodian, 2017). One consequence of adolescent brain development is the dramatic improvement in processing abilities. Unlike young children, who take in a limited number of highly frequent, meaningful and predictable language categories (De Ruiter et al., 2018), older learners can process linguistically complex texts, given their vastly improved cognitive capacities of attention, memory, and processing (Kilford, Garrett, & Blakemore, 2016; Larsen & Luna, 2018). During adolescence, multimodal integration areas that support higher order cognition undergo structural and functional maturation, enhancing connectivity and task-induced activation (Simmonds et al., 2014), including executive control functions (Crone, 2009). Thus, development increases the complexity of the human learning architecture, allowing learners to filter their attention to less frequent and prominent features in linguistic systems (Ramscar et al., 2018). To accomplish this, the learning system itself changes with time and experience to adapt to the changing structure of the linguistic input (Onnis & Thiessen, 2013), such as the shift from the spoken to the written modality (Berman, 2008).

The impact of these patterns of reorganization in the maturing brain on language learning was construed by Karmiloff-Smith (1992) as reflecting a process whereby knowledge becomes more integrated, denser, and more readily accessible in adolescence (Tibi, Tock, & Kirby, 2019). Along similar lines, Ramscar and Glitcho (2007) show that language acquisition changes its character from unsupervised learning in early childhood to the more agentive, self-monitored, supervised learning typical of older childhood and adolescence. These developments support enhanced metalinguistic abilities and access to higher-order, nonliteral language (Berman & Ravid, 2010; Karmiloff-Smith 1992). They also enable increasing text-production abilities (Berman, 2008; Nippold & Scott, 2009).

Later language development is thus characterized by the emergence of mature, complex language capabilities (Berman, 2017; Clark, 2004; Kuhn, 2011; Proverbio & Zani, 2005). Language skills continue to develop across the school years until young adulthood and beyond (Berman, 2007; Berninger et al., 2017; Nippold, 2016). Brain and cognitive developments enable older language learners to extract information from larger and more diverse samples of the data and learn less prominent categories and items. A critical part of the immense growth in language knowledge and skills is due to the consolidation of linguistic literacy during the school years (Egbert, 2020; Ravid & Tolchinsky, 2002), including the ability to shift flexibly between the spoken and written modalities, produce monologic discourse in different text types, adopt different perspectives on events and situations, and gain command of figurative language (Ashkenazi & Ravid, 1998; Beltrán-Planques & Querol-Julián, 2018; Berman, 2008; Berman & Ravid, 2010; Brandes & Ravid, 2019; Colston, 2020; Vulchanova et al., 2019).

A critical component of language proficiency attainment towards adulthood concerns lexical development in the content word domain of nouns, adjectives, and verbs (Beitchman et al., 2008; Clark, 2017). Maturing cognitive and interpersonal skills and the consolidation of linguistic literacy usher in abstract reasoning and increasing analytic capability (Crone, 2009; Fortman, 2003), which find expression in complex words typical of written, academic language (Anglin, 1993). In a recent article, Atanasova et al. (2020) summarize a large body of data showing that the difference between children's and adults' lexical knowledge, measured in terms of accuracy and speed ,is both quantitative and qualitative. Their study indicates that with regards to AoA of words in production, young adolescents (14- to 16-year-olds) display intermediate behavior between younger children and adults that may indicate ongoing brain maturation. Studies on Hebrew, where the lexicon is organized by morphological devices and systems (Ravid, 2019), show that command of derivational morphology and morpho-lexical abilities increases exponentially across adolescence to include mental words, abstract and derived nouns (Ravid, 2006; Ravid & Avidor, 1998), denominal adjectives (Cutillas & Tolchinsky, 2017; Ravid et al., 2016), and specialized vocabulary (Ben Zvi & Levie, 2016; Berman & Nir-Sagiv, 2010; Osterhaus, Koerber, & Sodian, 2017). Lexical growth in adolescence is accompanied by a dramatic increase in syntactic complexity, which is mainly used to identify, comprehend and express complex ideas in texts and to organize the flow of information in a text (Bybee & Noonan 2001; Ravid and Berman, 2006).

Writing in language development

Writing is prototypically a pre-planned, non-interactive activity, impersonal and nondirect - a process that results in editable text (Hyes & Flower, 1981; Murray, 2012). The generation of stable textual products disengaged from their immediate context of production constitutes the basis for a literate society that documents and comments upon knowledge (Olson, 2006). It also brings to consciousness the structures, semantics and functions of linguistic usage that may be absent from awareness in oral expression (Olson, 1994). Writing is considered the ultimate achievement of linguistic literacy (Berman & Ravid, 2008; Ravid & Tolchinsky, 2002), imposing cognitive demands on memory, executive functions, and top-down processing, and promoting the creation of complex syntactic units (Chafe, 1994; Kärkkäinen et al., 2007; Slobin, 2003).

In an educational context, writing is the highway to learning, processing, organizing, storing and retrieving information during the school years, as well as for communicating with teachers and peers (Donovan, 2001). The older the students and the higher their grade level, the more important the quality of writing and its quantity become in integrating information from external resources. Therefore, writing activities increasingly occupy the central arena of linguistic abilities in school age children, while at the same time challenging them with a paradoxical demand. On the one hand, writing calls for the creative production of new content in line with the modality, the communicative circumstances, and the required genre; but at the same time, constructing a piece of written discourse imposes heavy demands on both bottom-up and top-down processing abilities. Writers need to retrieve the specific words for the desired expression of content, combine them in the appropriate syntactic and rhetorical structures, and integrate them smoothly and meaningfully in view of the overall goal of the text under construction, paying attention to facets of the notational system such as spelling and punctuation (McNamara, Crossley, & McCarthy, 2010). To achieve this balance in expressing communicative content in a constrained form, students need to access rich cognitive resources, such as monitoring and executive control, meta-memory, planning, setting goals, and manipulating series of units.

It is no wonder that gaining command of writing abilities is a protracted process requiring both internal resources as well as active mediation, support and guidance by expert teachers. This process interfaces with later language development, as described above, across the school years (Berman, 2005; Nippold, 2016; Silliman & Berninger, 2011). Importantly, it depends on linguistic resources that consolidate over the school years (Berman, 2014; Castillo & Tolchinsky, 2018; Graham & Harris, 2009), in tandem with socio-cognitive developments that take place in adolescence (Blakemore, 2012). A series of psycholinguistic studies compared morpho-syntactic, lexical, and discursive features of written and spoken Hebrew texts produced by native, non-expert school-aged and adult writers. Written texts were found to be informationally denser, lexically richer and more diverse, containing more high-register lexical items and morphological constructions than spoken texts (Berman & Nir-Sagiv, 2010; Ravid & Berman, 2009). Written texts were also shown to have fewer discourse markers, repetitions, false starts, hedges, and other disfluencies than spoken texts (Ravid & Berman, 2006). Written more than spoken texts contained abstract and morphologically complex nominals (Ravid, 2006; Ravid & Cahana-Amitay, 2005), often modified by derived adjectives in the attributive position (Ravid & Levie, 2010), as well as longer and more complex noun phrases (Ravid & Berman, 2010; Ravid et al., 2002), often in the form of heavy compounds (Ravid & Zilberbuch, 2003). These linguistic features increased with age and schooling levels (Berman, 2004, 2007; Berman & Nir-Sagiv, 2007; Ravid, 2006).

In addition to the *spoken – written* dichotomy, the factor of *genre* knowledge profoundly affects all linguistic and discursive domains (Louwerse, McCarthy, McNamara & Graesser, 2004; Figueroa, Meneses & Chandia, 2018; Snow & Uccelli, 2009). We know that in addition to reader competencies that depend on text genre (Best, Floyd & McNamara, 2008), text production abilities are also highly dependent on text genre (Berman & Nir-Sagiv, 2009; Brandes & Ravid, 2016, 2019). A plethora of developmental psycholinguistic studies have examined narrative versus expository writing across the school years (Berman, 2008; Berman & Katzenberger, 2004; Nippold & Scott, 2010; Nippold & Sun, 2010; Ravid & Berman, 2010), focusing on text content, discourse structure, lexicon, and morphosyntax. These studies indicate that growing familiarity with different text types and genres enables writers to employ various genre-appropriate complex syntactic constructions by adolescence (Aparici, Rosado, & Perera, 2016; Brimo & Hall-Mills, 2019; Ravid, 2005, 2013; Ravid, Dromi & Kotler, 2009). The analysis of narrative and expository texts has also served to highlight differences in the syntactic skills of disordered or deprived and typically developing children and adolescents (Berman, Nayditz & Ravid, 2011; Gillam & Johnston, 1992; Nippold et al., 2008, 2009; Scott & Windsor, 2000).

Academic writing

The current study examines the linguistic constructions typical of the non-narrative continuum in academic writing. At one end of this continuum lie expository texts, characterized by argumentation or persuasion regarding social / conceptual issues; and at the other lie informative texts, describing factual (or less controversial) phenomena. The previous section shows that expository texts have been the focus of many developmental psycholinguistic studies, especially contrasting them with narrative writing. However, while informative texts (together with expositories) constitute the prevalent genre in academic reading and writing, to date, not much research has been carried out on students' developing linguistic and discursive expressive abilities in this genre (Beers & Nagy, 2011). This is especially necessary in view of the fact that writing expository and informative texts is commonly requested in school (Donovan & Smolkin, 2001; Nippold et al., 2009; Schleppegrell, 2003). Therefore, expository and informative texts constitute the genres of the current study on the development of academic writing in school age. The characteristics and the specific communicative purposes of each genre, as elaborated below, are expected to affect students' production.

Table 1: The main properties of expository and informative texts: Typical Stance, Content, Protagonists.

Genre → Properties	**Expository**	**Informative**
Stance	Abstract, generic	Concrete, generic
Content	Presentation of socio-cultural issues and concepts	Description of objects, people, natural and man-made events, processes
Protagonists	Ideas	Informational topic

Table 1 depicts the main features of expository versus informative texts in terms of Stance or tone, content, and the typical protagonists (Berman, 2005; Ravid, 2005; Scott & Balthazar, 2010). The language and general tone or stance of expository texts (on themes such as *animal consciousness*, *how to eliminate violence*) is abstract and generic, with ideas as the main protagonists and socio-cultural issues as the typical content. The literature shows that from the early school years, expository texts, which depict the unfolding of ideas, are highly nominal and contain high register vocabulary, including lexically diverse, abstract nouns, and denominal adjectives. They are characterized by complex syntax (both subordinating and coordinating), heavy NPs, non-dynamic constructions, passive voice, non-finite verbs, and epistemic hypothetical constructions. Thus, it takes the whole span of adolescence to achieve writing skills in expositories (Cutillas & Tolchinsky, 2017; Nippold & Sun, 2010; Oblinghouse & Wilson, 2013; Ravid & Berman, 2010; Ravid & Levie, 2010).

In contrast, informative texts (e.g., *how lightning works*, *perfumes*) focus on description of objects, events and natural processes from a concrete yet generic point of view. They present information about the world in ways that are intrinsic to these topics. We are in possession of less knowledge about the development of informative text production – certainly much less than what we know about narrative and expository text production – and that is despite the fact that they constitute the majority of texts read and written across the school ages. Research so far has focused on lower grades (Tower, 2003), showing that 5th graders used more basic structures in writing informative texts versus more sophisticated structures in narrative texts (Donovan, 2001). On the other hand, students' informative texts contained more content words, more complex syntax and elaboration than narrative or persuasive texts (Beers & Nagy, 2011; Olinghouse & Wilson, 2013). And as for Hebrew, a single study on school-going populations shows that Hebrew-speaking 7th graders wrote very concrete, scarcely developed informative texts (Ravid & Shalom, 2012).

In sum, investigating the texts that school-goers construct provides a window on their linguistic abilities, in a period when command of written language is opening up new avenues to linguistic knowledge and on the genre-specific construction of different text types (Jisa & Tolchinsky, 2009; Sanders & Schiperoord, 2006).

Methods

Against this background, the current study is based on the assumption that gaining command of the structures and functions of texts of different genres is one of the main goals of linguistic literacy in education. The study aimed to gain psycholinguistic information and insights regarding the development of informative versus expository text writing abilities in Hebrew-speaking elementary, middle and high school students, compared with adults. Our study focuses on features such as text content, text structure, lexicon, and morpho-syntax. We aimed to determine when and to what extent participants are able to provide a full description of an informational topic that is familiar to them, and to explore the linguistic and cognitive resources they employ in elaborating on an informational topic – and thus to shed light on the language-cognition interface in writing development (Baaijen & Galbraith, 2018; Deane et al. 2008; Kellogg, 2008; Sanders & Schiperoord, 2006). A related aim of this study was to specify the characteristics of expository versus informative texts, two genres that constitute two ends of the non-narrative continuum. In practical terms, we aimed to (i) assess students' written products in the context of their age and literacy level; and to (ii) pinpoint certain areas where some of the students might be in need of remediation (Berman et al., 2011; Scott & Balthazar, 2010; Scott & Windsor, 2000) - as against the two main variables of the study – genre (expository vs. informative text type) and age / schooling, as described below.

Hypotheses

We had two main hypotheses. The first one was that all textual measures (see below) would increase with age and schooling (Berman & Katzenberger, 2004; Crossley et al., 2011; Nippold, Ward-Lonergan, & Fanning, 2005). A second hypothesis predicted differences between the expository and informative texts (Tolchinsky, 2019). However, we could not predict the direction of this hypothesis, as this is a first study of its kind in Hebrew.

Participants and data base

Participants constituted 547 students (roughly half male, half female) in three school levels, and adults, as follows: 139 4^{th} graders, 135 7^{th} graders, 126 11^{th} graders, and 147 educated adults with post-high school education. All participants were monolingual, native Hebrew speakers, with typical development, from mid-high SES schools. The school children and adolescents were recruited in their own classrooms on a voluntary basis. Given the educational and psycholinguistic nature of the current study, no students were excluded from participation, as we wished to represent the whole range of grade-level writing performance in typical schools of the Israeli national system (Bar-On & Ravid, 2011).

Each participant wrote two texts, as delineated in Table 2: one informative text and one expository text, yielding a data base of 1,094 texts. Students wrote both texts (in the order they chose) within a school period of 45 minutes, the length of a typical lesson. Students who asked for more time were granted an extension until they were done. Adults were recruited individually, and having filled out an internet questionnaire about their background, they each wrote the requested texts and sent them to the first author.

The topic of the expository text was Success and Failure, and the instructions for participants were as follows: *"Success and failure are topics which interest youth and adults, and every person has different opinions and understandings regarding these topics. Think about the topic of success and failure, about their reasons and outcomes, and write an exposition that will present your thoughts on the topic"*. The topic of the informative text was The Cellular Phone, and the instructions for participants were as follows: *"Write a text that presents and describes the cellular phone (for example, the Iphone or the Android). Do not write a story: describe the cellular phone, its uses and functions, as though you were writing for Wikipedia. Also think of the implications of the use of the cellular phone for people and for society"*.

Table 2: The study database (N = 1,094 texts).

Informative texts	**Expository texts**
Cellular phones	Success and failure
N = 547	N = 547

Coding and analyses

Two kinds of dependent variables were coded in the texts. A first type were *count* variables that measured the volume of the text in three different ways, as follows: (i) number of words; (ii) number of clauses; and (iii) Mean Clause Length (MCL), the number of words divided by the number of clauses, a derived measure of lexical and syntactic density in the clause (Berman & Ravid, 2008; Ravid, 2005). Text size as assessed by these three measures gives us a good idea of students' productive ability to express themselves in writing. A second type of variable, related to the linguistic properties of the texts, was *recompensed* – variables designed to reflect the linguistic properties of the text, adjusted for number of words, i.e., calculated as a ratio of variable/length in words. Recompensed variables were lexical (abstract nouns, attributive adjectives), morpho-syntactic (compounding devices), syntactic (conjunct structures, complement clauses, headless clauses, and relative clauses), as well as demarcation markers (a measure composed of connecting words and punctuation marks). These linguistic features were found in the literature described above to be typical of the literate language typical of advanced, though non-expert, writing. The Results and Discussion sections will elaborate on these structures and their meaning for the development of writing skills in later childhood and adolescence.

Analyses of both count and recompensed dependent variables were carried out with regards to the two key variables in the current study – *age/schooling* - a developmental perspective, with the adults as a point of comparison given their mature language abilities; and *genre*, with the idea that text type affects the linguistic properties of the text. We also expected age/schooling differences to be expressed differently in the two text types. Analysis was carried out in two stages for each textual measure, using the Generalized Estimating Equations (GEE) model. For each analysis, the first stage started with genre and age/schooling level, and then examined the interaction of these two variables. Interactions were analyzed within each text type using the Bonferroni Pairwise test in contrasting every possible age/schooling pair.

Results I: Text size - count variables

We start by presenting the results of the text size measures – words, clauses and MCL – in terms of the two variables of age/schooling and genre. Appendix A (i, ii, iii) shows the results of the GEE model for these three analyses. Table 3 summarizes the results of the GEE model. Appendix A (i) and Table 3 indicate that texts written by adults were longer in words than those written by school children, but shorter than those written by 11th graders, with no effect for genre, but with an interaction that is shown in Figure 1.

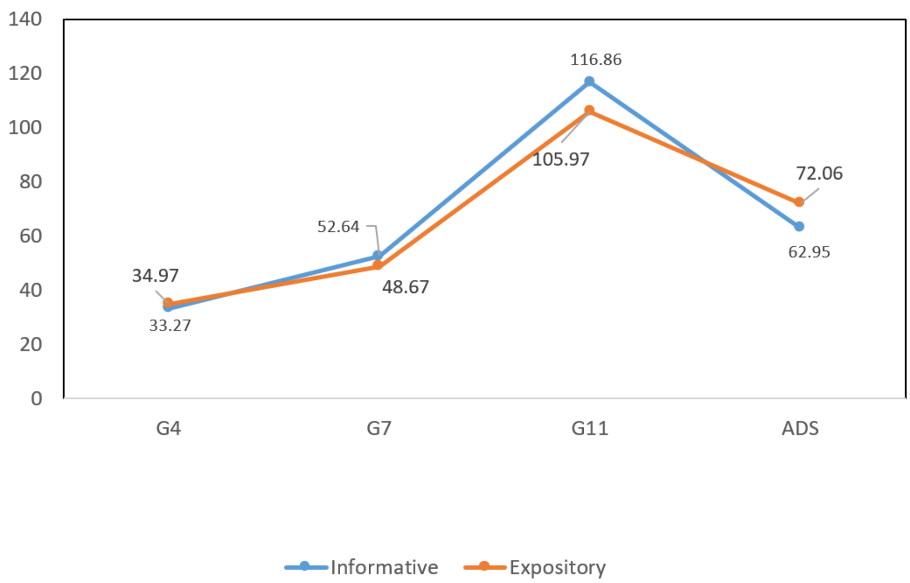


Figure 1: Interaction of age/schooling and genre in mean number of words.

Appendix A (ii) and Table 4 indicate that texts written by adults were longer in clauses than those written by younger children, but shorter than those written by 11th graders; and that expository texts had more clauses than informative texts. The interaction is shown in Figure 2.

Table 3: Summary of differences between age groups and genres regarding mean number of words, clauses, and Mean Clause Length.

Category	Age/Schooling	Genre	Interaction
# Words	Ads > 4th (b=0.51, p<.001), Ads > 7th (b=0.75, p<.001); 11th > Ads (b=1.62, p<.00)	X	✓
# Clauses	Ads > 4th (b=0.78, p<.001); 11th > Ads (b=1.89, p<.001)	E > I	✓
Mean Clause Length	Ads > 4th, 7th, 11th (b=-2.01,-1.20,-.077, p<.001)	X	✓

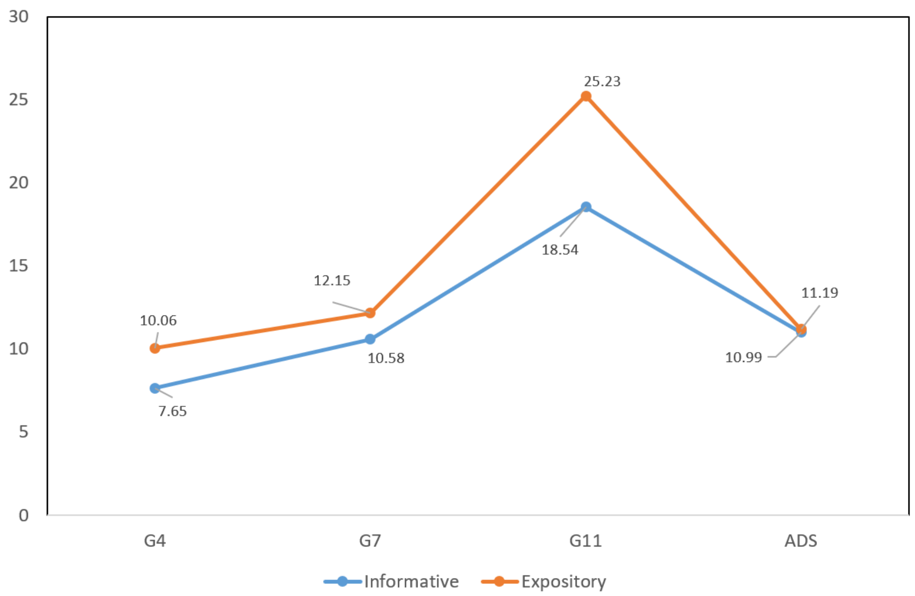


Figure 2: Interaction of age/schooling and genre in mean number of clauses.

Appendix A (iii) and Table 3 indicate that adult clauses were longer than clauses in texts by younger writers; and informative clauses were longer than expository clauses. Figure 3 shows the interaction.

Summary of results regarding text size

Our first hypothesis was partially confirmed. 11th graders produced the most words and clauses, and 4th graders produced the fewest numbers of words and clauses. MCL increased with age, with the longest clauses in the adult texts. Although the number of words did not differ across genres, expository texts had more clauses, and MCL was larger in the informative texts.

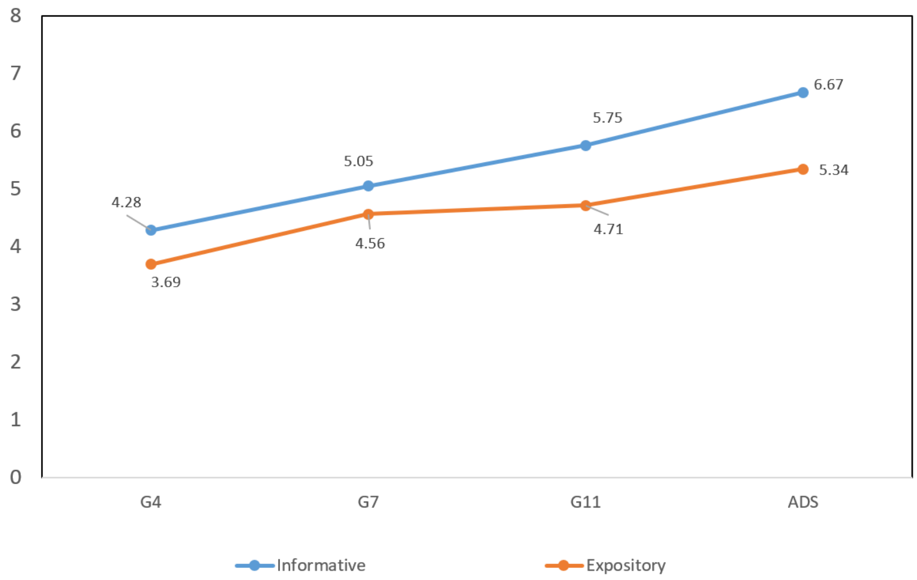


Figure 3: Interaction of age/schooling and genre in mean clause length (MCL) – words divided by clauses.

Results II: Recompensed lexical, morpho-syntactic and demarcation variables

We now move on to the linguistic measures characterizing our participants' texts, which were recompensed by standardizing text length. A proportion of each variable was calculated in relation to the number of words in the text, given that text sizes are different in the different age/ schooling groups (Salas & Caravolas, 2019).

Lexical measures

The lexical measures focused on two classes of words in Hebrew, which have been shown to be of critical importance in later language development: abstract nouns (e.g., *xoxma* 'wisdom', *kabala* 'reception') and attributive, specifically denominal adjectives (e.g., *merkazi* 'central' from *merkaz* 'center') (Ravid, 2006; Ravid & Levie, 2010; Ravid et al., 2016).

Abstract nouns. **Abstract nouns** were identified in the texts and analyzed in two different ways: their frequency of usage (**tokens**); and their category size, i.e., the number of different abstract nouns (**types**). In addition, **abstract nouns in compounds** (e.g., *havtaxat ha-eyxut* 'assurance (of) the-quality') were identified and analyzed both as **tokens** and **types**, as indicating the production of heavy noun phrases with abstract cores (Ravid & Berman, 2010). Appendix B (i) shows the results of the GEE model for the recompensed abstract noun analyses. Table 4 summarizes the results of the GEE model. Figures 4-5 present the interactions.

Appendix B (i), Figures 4-5 and Table 4 indicate that the cutoff point in development re-

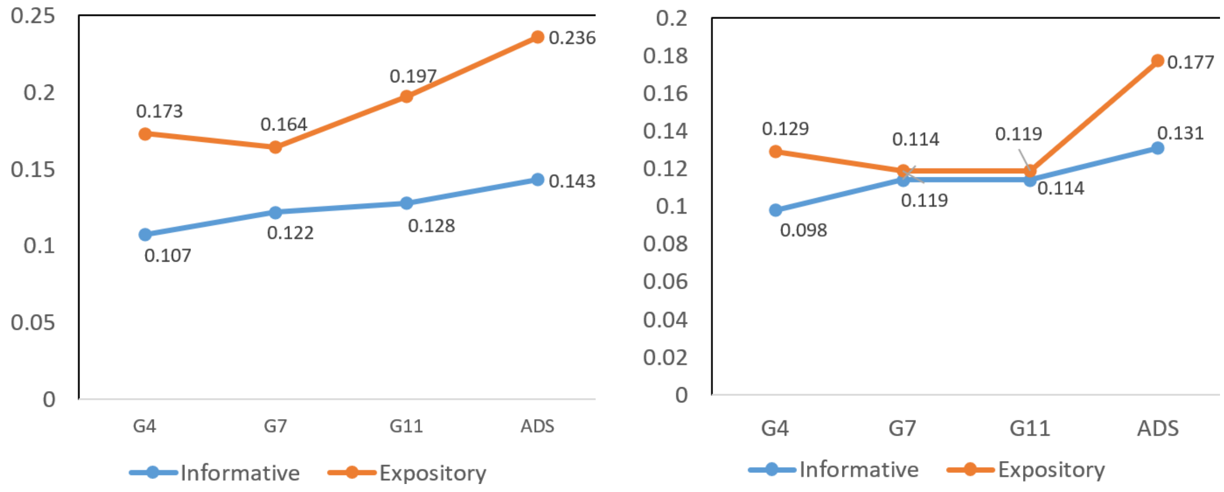


Figure 4: The age/schooling and genre interactions on numbers of recompensed abstract nouns in the texts (tokens on the left, types on the right).

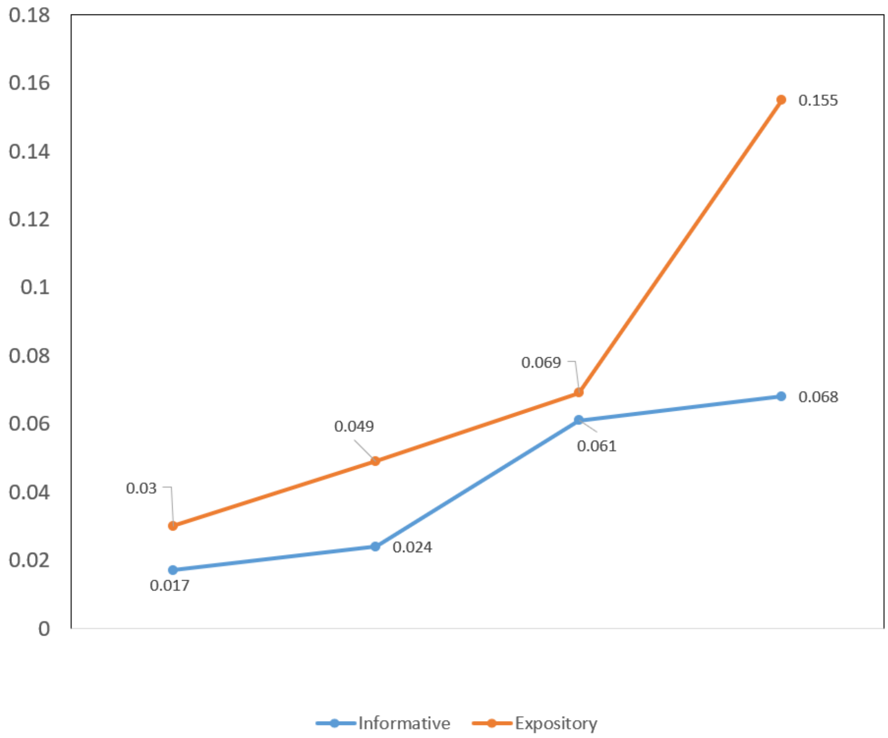


Figure 5: The age/schooling and genre interaction on compounds with recompensed abstract noun types in the texts.

Table 4: Summary of differences between age groups and genres regarding recompensed abstract noun measures.

Category	Age/Schooling	Genre	Interaction
Abstract noun tokens	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-.30, -.26, -.15, *p*<.001)	E > I	✓
Abstract noun tokens in compounds	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-1.41, -1.05, -.37, *p*<.001)	E > I	X
Abstract noun types	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-.30, -.26, -.26, *p*<.001)	E > I	✓
Abstract noun types in compounds	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-.30, -.26, -.26, *p*<.001)	E > I	✓

garding abstract nouns is in later adolescence – on all analyses, adults have more abstract nouns than the younger groups; and expository texts are richer in abstract nouns than informative texts.

Attributive adjectives. **Attributive adjectives** were identified in the texts and analyzed in two different ways: their frequency of usage (**tokens**); and their category size, i.e., the number of different attributive adjectives (**types**). In addition, a specific class of attributive adjectives – **denominal adjectives** (e.g., *eyxuti* 'of high quality' from *eyxut* 'quality') were identified and analyzed both as tokens and types, as indicating a literate lexicon (Ravid & Levie, 2010; Ravid et al., 2016). Appendix B (ii) shows the results of the GEE model for the analysis of recompensed attributive adjective analyses. Table 5 summarizes the results of the GEE model regarding attributive adjective analyses. Figure 6 presents the interactions.

Table 5: Summary of differences between age groups and genres regarding recompensed attributive adjective measures.

Category	Age/Schooling	Genre	Interaction
Attributive adjective tokens	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-.39, -.35, -.18, *p*<.001,.001,.01)	I > E	X
Attributive adjective types	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-.43, -.38, -.22 , *p*<.001)	I > E	X
Denominal attributive adjective tokens	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-1.77, -1.09, -.54 , *p*<.001)	I > E	✓
Denominal attributive adjective types	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-1.78, -1.16, -.70 , *p*<.001)	I > E	✓

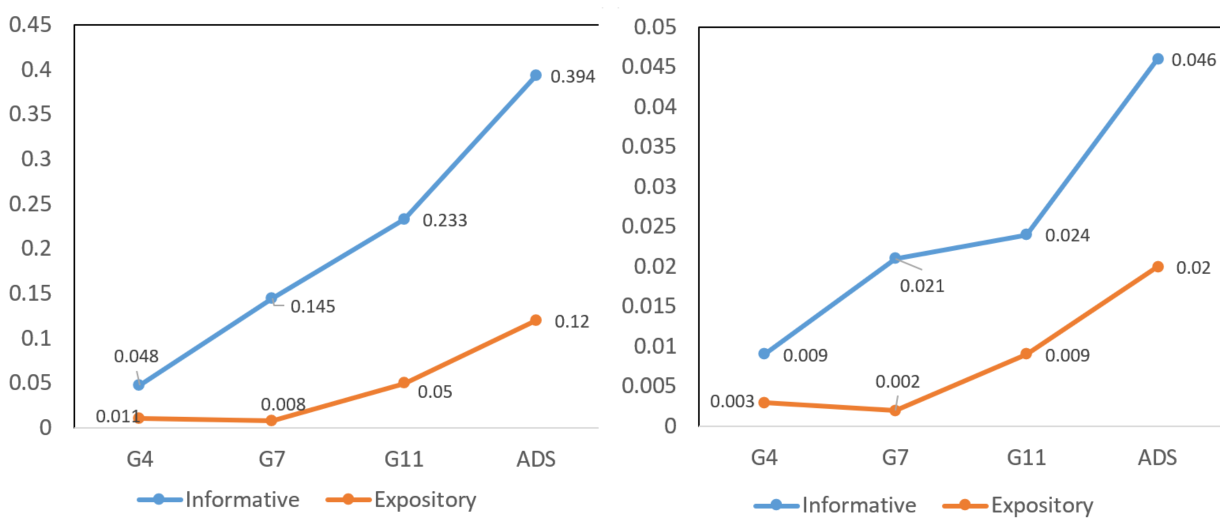


Figure 6: The age/schooling and genre interactions on recompensed denominal attributive adjectives in the texts (tokens on the left, types on the right).

Summary of lexical measures

Our first hypothesis was fully confirmed: development had an effect on all lexical measures, which increased with age and schooling and were significantly more abundant in the adult group. Moreover, as per our second hypothesis, genre plays an important role in learning to produce academic texts: while abstract nouns are always more numerous in expository texts, all analyses of attributive adjectives, including denominal adjectives which are the hallmark of Hebrew literacy (for example, *ta'asiyati* 'industrial' from *ta'asiya* 'industry'), indicate they are more numerous in the informative texts. Note that this does not imply that informative texts are more "academic" than expositions, but rather more linked to school-type writing, as elaborated in the Discussion.

Morpho-syntactic measures

The morpho-syntactic measures focused on compounds of various kinds in Hebrew, which straddle the boundary between morphology and syntax – and have also been shown to be of critical importance in later language development, given their role in enhancing the structure and semantics of noun phrases (Ravid & Zilberbuch, 2003; Ravid & Assuline Tsabar, 2017). Three compound structures were coded and analyzed in terms of tokens: Classical bound adjacency compounds (e.g., *anfeŷha-ec* 'brancheŝthe-tree', where the compound head is morphologically bound to its complement; an English example is network reception); **free compounds** (*ha-anafim shel ha-ec* 'the-branches of the-tree', or *ha-klita shel ha-reshet* 'the reception of the network'), considered to be the gateway to heavy nominal syntax (Ravid & Assuline Tsabar, 2017); and **chained compounds** (e.g., *masax ha-maga shel ha-maxshir* 'the touch screen of the device'), which are composed of chains of compounds.

Appendix B (iii) shows the results of the GEE model for the analysis of recompensed compounds. Table 6 summarizes the results of the GEE model regarding compound analyses. Figure 7 presents an interaction.

Table 6: Summary of differences between age groups and genres regarding recompensed compound measures.

Category	**Age/Schooling**	**Genre**	**Interaction**
Adjacency compound tokens	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-.99, -.64, -.18, *p*<.001,.001)	I > E	X
Free compound tokens	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-.79, -.54, *p*<.001)	X	✓
Chained compound tokens	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=-1.34, -1.36, -.62, *p*<.001)	I > E	X

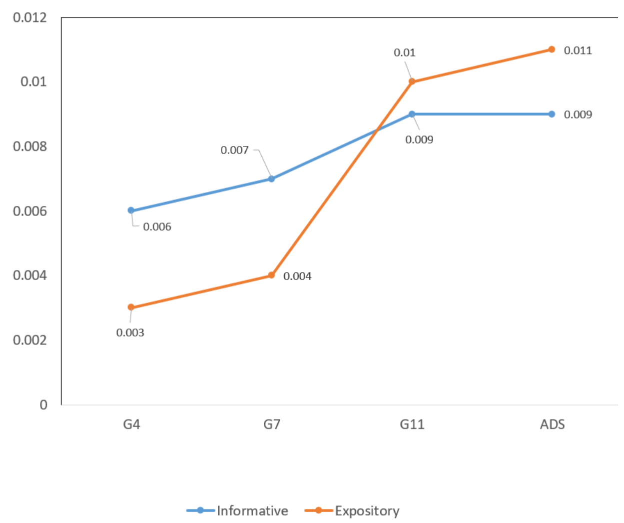


Figure 7: The age/schooling and genre interactions on recompensed free compounds in the texts.

Summary of morpho-syntactic measures

Our first hypothesis was fully confirmed: development had an effect on all compound measures, whose numbers increased with age and schooling and were significantly more abundant in the adult group. Moreover, as per our second hypothesis, genre plays an important role in learning to produce academic texts: Adjacency compound tokens and chained structures are more numerous in the informative texts – again, as elaborated in the Discussion, as a measure that reflects the facilitation of school-type language by this genre.

Syntactic measures

This analysis focused on two kinds of categories – syntactic categories and demarcation markers. Appendix C shows the results of the GEE model for the analysis of these recompensed structures. Tables 7 and 8 summarize the results of the GEE model regarding these analyses. Figures 8 and 9 present the interactions.

Syntactic structures. The syntactic measures focused on four categories that have been shown to characterize written text production in Hebrew-speaking higher grade levels and adults. First, **conjunct structures**, coordinating structures attached to a single unit as in *success is critical in life and in school*, where *critical* governs *in life* and *in school* (Ravid, 2013; Ravid & Hershkovitz, 2017); **complement clauses**, typically composed of a mental or dicendi verb followed by a clause delineating the content of the mental of saying activity, e.g., *he explained that failure might occur*; **headless relative clauses**, e.g., *someone who received a low grade*; and **relative clauses**, e.g., *a device that changed our lives.* While relative clauses are known to characterize richer academic language (Brandes & Ravid, 2016; Ravid & Berman, 2010: van Rijt, van den Broek, & De Maeyer, 2021), this is the first time we examine the frequencies of headless relatives and complement clauses in the development of text production. Appendix C (i, ii, iii, iv) and Table 7 provide the statistical analyses and summary of differences for the syntactic measures. Figure 8 shows the interactions.

Table 7: Summary of differences between age groups and genres regarding recompensed attributive adjective measures.

Category	**Age/Schooling**	**Genre**	**Interaction**
Conjunct structures	Ads > 4^{th} (*b*=-.21, *p*=.001)	E > I	✓
Complement clauses	Ads > 4^{th}, 7^{th} (*b*=-1.30,-.76 , *p*<.05)	E > I	X
Headless clauses	Ads > 4^{th}, 7^{th}, 11^{th} (*b*=.92,.49,.29 , *p*<.001,.01,.05)	E > I	X
Relative clauses	Ads > 4^{th}, 7^{th} (*b*=-.79, -.45, , *p*=.001)	I > E	✓

Appendix C (i-iv), Table 7 and Figure 8 indicate that all syntactic measures increased with age and schooling, with the cutoff point between 7^{th} or 11^{th} grade and adults; and that genre had an effect on these productions, as conjunct structures, complement structures and headless relatives were more numerous in expository than informative texts; whereas relative clauses were more abundant in informative than expository texts.

Demarcation markers. Alongside the syntactic measures we examined **textual demarcation**, a measure of information flow that takes into account connectivity and coherence in writing. This measure took the form of three analyses: **Conjoining lexical markers**, such as *in addition, first and foremost*; **commas** and **full stops**. All measures were recompensed by number of words. Appendix C (v, vi, vii) and Table 8 provide the statistical analyses and summary of differences for the syntactic measures. There were no interactions.

Appendix C (v, vi, vii) and Table 8 indicate that text connectivity and demarcation improved with age and schooling as expressed by the three measures presented above; and

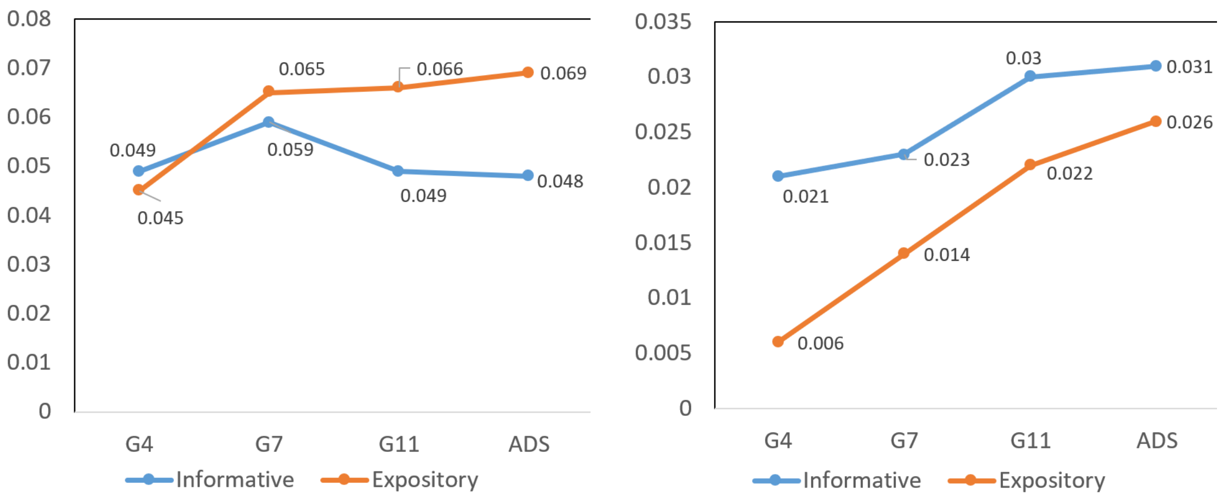


Figure 8: The age/schooling and genre interactions on recompensed conjunct structures in the texts (on the left) and relative clauses (on the right).

that informative texts were better demarcated than expository texts.

Table 8: Summary of differences between age groups and genres regarding recompensed demarcation measures.

Category	**Age/Schooling**	**Genre**	**Interaction**
Connective lexicon	Ads > $4t^{th}$, 7^{th} (b=.20, .27, p~.05, <.01)	I > E	X
Commas	Ads > 4^{th}, 7^{th} (b=-1.23, .-26, p<.001, <.05)	I > E	X
Full stops	Ads > 4^{th}, 7^{th} (b=-.39, .-16, p<.001, ~.05)	X	X

Summary of syntactic and demarcation measures

Our first hypothesis was fully confirmed: development had an effect on all syntactic and demarcation measures, whose numbers increased with age and schooling and were significantly more abundant in the adult group. Moreover, as per our second hypothesis, genre plays an important role in learning to produce academic texts: most syntactic structures were more numerous in expository texts (whereas relative clauses were more numerous in informative texts); and informative texts were better demarcated than expository texts in terms of connectives and commas.

Discussion

The current study investigated the linguistic characteristics of 1,094 informative and expository texts written by 547 native Hebrew-speaking participants in three grade levels (4^{th}, 7^{th}, and 11^{th} grades), compared with educated (but non-expert) adults. Informative texts described cellular phones, while expositions discussed the themes of success and failure. Each text was coded and analyzed by lexical, morphological, and syntactic measures.

In line with the literature reviewed above, we had two major hypotheses. One expected to detect more and richer linguistic measures with age and schooling. A second hypothesis expected to see differences in the distributions of these measures across the two genres, but since informative texts have been scarcely investigated in the framework of developmental text production, we did not have a direction for these expected differences. Indeed, both our hypotheses were confirmed in interesting ways. Below we discuss the results in two perspectives – a developmental outlook and a genre-specific outlook.

A developmental perspective on written text production

A broad array of studies describe the pathway undertaken by children and adolescents in developing their writing skills across the school years. A critical part of the immense growth in language knowledge and skills during these years is due to the consolidation of linguistic literacy (Egbert, 2020; Ravid & Tolchinsky, 2002), including the ability to shift flexibly between the spoken and written modalities, produce monologic discourse in different text types, adopt different perspectives on events and situations, and gain command of figurative language (Ashkenazi & Ravid, 1998; Beltrán-Planques & Querol-Julián, 2018; Brandes & Ravid, 2019; Colston, 2020; Vulchanova et al., 2019).

In Hebrew, special attention has been paid to the developing linguistic features of text production abilities across the school years, such as high-register lexicon (Ravid, 2004; Ravid & Berman, 2009); optional inflectional morphology (Cahana-Amitay & Ravid, 2000); derived, abstract nouns (Ravid, 2006; Ravid & Avidor, 1998; Ravid & Cahana-Amitay, 2005); compounding structures (Ravid & Zilberbuch, 2003); denominal and syntactically expanded adjectives (Berman, Naydic & Ravid, 2011; Ravid & Levie, 2010: Ravid et al., 2016); large noun phrases (Ravid & Berman, 2010; Ravid et al., 2002); prepositional phrases (Brandes & Ravid, 2016); adverbial clauses (Brandes & Ravid, 2019); conjunct structures (Ravid & Hershkovitz, 2017); and a plethora of other morpho-syntactic, discursive and content measures (Berman, 2018; Ravid, Dromi, & Kotler, 2009). Across all of these domains in diverse studies, the high-register, abstract, lexically specific, alternative linguistic devices characterizing richer, advanced language increased with age and schooling, especially in later adolescence and in adulthood as compared to younger writers.

The current findings of morpho-syntactic analyses in a new corpus of over 1,000 written texts support and enhance this picture of the increasing linguistic complexity in the development of writing. While 11^{th}graders wrote the longest texts in raw numbers of words and clauses, the derived measure of MCL, which provides mean clause length in words, showed that clauses increased in length with age and schooling, and that adults had the longest clauses. Clauses (simplex sentences) can become longer in two ways – either by having more phrases (nominal, prepositional), which indicate more argument or circumstance structures; or by having longer phrases, with more noun and verb modifications. In both ways, longer clauses convey more information in clause-internal, syntactically complex structures. Thus, text production involves selecting your wording carefully with fewer repetitions and learning to produce more meaningful, tightly packaged clauses with age and schooling (Elsabbagh & Karmiloff-Smith, 2004; Northey, McCutchen, & Sanders, 2016). For example, a single clause describing the cellular phone in an adult text reads as follows (loosely translated from Hebrew) *cellular telephone devices in our times are usually rectangular, hand-sized, with several buttons.* Writing such clauses is part of the literacy and language skills needed from linguistically proficient individuals, aided by cognitive, linguistic and literacy development, who are able to put together a coherent text consisting of multiple ideas (Crossley et al., 2011; Greg & Nelson, 2018; Lindgren, Leijten, & Van Waes, 2011; McNamara et al., 2010).

More support for this developmental view was found in the lexical, morpho-syntactic and syntactic analyses performed while taking text length in words into consideration. As the results section indicates, the recompensed numbers of abstract nouns (e.g., *madad* 'measure', *mexkar* 'research') increased with age and schooling, culminating in a dramatic rise in adults; and they also appeared more in complex NPs based on N-N compounding, e.g., the high-register 'double' compound *haclaxato shel adam* 'lit. the-success,3rd.Sg.Masc of (a) person', or the chained compound *yexolet hafacat ha-meyda* '(the) ability (of) dissemination (of) information'. According to Schmid (2012), many abstract nouns act as conceptual shells that encode propositions and larger information chunks within a noun phrase, and serve as cohesive devices as their content is determined by referring to their context – very typical of academic discourse, of which both expository and informative texts are examples (Prados, 2018). It is no wonder that in addition to amassing the discipline-related and world knowledge lexicon that is necessary to expressing abstract concepts in academic writing (Khokhlova, 2014), it takes developmental time to gain command of encasing them correctly in dense syntactic environments (Nippold et al., 1999; Ravid, 2006). Attributive adjectives typically qualify nouns, e.g., *shipurim xadashim* 'new improvements'. They are thus optional nominal modifiers whose main function is qualification - that is, to subdivide classes or specify among particular instances within classes on the basis of characteristics (Bolinger, 1967; Feris, 2014; Nelson, 1976) – which are among the properties of grammatically complex and mature linguistic expression (Biber, Gray & Staples, 2016; Parkinson & Musgrave, 2014; Staples et al., 2016). Just like abstract nouns, the amounts of attributive adjectives rose with age and schooling, culminating in the adult group. This was also true of a specific morphological class of adjectives – denominal (noun-derived) adjectives such as *koli* 'voice-based' from *kol* 'voice', which are known to characterize literate, especially academic Hebrew (Ravid et al., 2016). Many of the N-A combinations in the older-group texts were composed of an abstract noun and a denominal attributive adjective, e.g., *zminut tmidit* 'constant accessibility' or *madadim xomriyim* 'concrete measures' – indicating a mature ability to describe an abstract entity with its specific qualification.

The domain of morpho-syntax was represented in this study by the class of compounds, which in Hebrew takes only the form of N – N combinations. Compounds are one of the major ways of expanding the Hebrew noun phrase (Ravid & Berman, 2010), especially in later language development (Ravid & Zilberbuch, 2003). The two major types of compounds – bound and free – again increased with age and schooling and were most numerous in the adult texts, e.g., *ma'agrey meida* 'information bases' or *migvan shel efsharuyot* '(an) array of opportunities'. These nominal combinations serve to express the complex subcategorization of entities that are the typical referents in academic texts. The chained compounds, where several nouns are chained together, depicting larger and complex entities, had the same distributional behavior, e.g., *ha-hashlaxot shel hamca'at ha-telefon ha-nayad* 'the implications of the invention of the cellular phone'.

Complex syntax is well known as an important property of the language of school aged children and adolescents (Brimo & Hall-Mills, 2019; Nelson, 2013), although much work has targeted younger children and / or populations with impairments (Balthazar & Scott, 2018; Delage, Stanford & Durrleman, 2021; Kawar, 2021; Schuele & Dykes, 2005). Lexical and morpho-syntactic devices - abstract nouns, denominal adjectives and compounding, as well as noun phrases - have been studied before in the framework of Hebrew text development research, as in other languages (see above). Also, clause packaging has been the highlight of several important developmental psycholinguistic studies by Berman and associates on text production in Hebrew-speaking school age children (Berman, 2014, 2018; Berman & Nir-Sagiv, 2009). However, this is the first time that syntactic coordinating and subordinating constructions are individually investigated to characterize the language of Hebrew academic texts in development.

We looked at four constructions, which all increased with age and schooling, and were significantly more prevalent in the adults' texts. Conjunct structures, as in *young children already* ***consolidate their self appreciation and self perception*** *from early on* (conjoined elements bolded), constitute a arena of much relevance to the development of writing in Hebrew. While being a coordinating rather than subordinating device, Ravid (2013) and Ravid and Hershkovitz (2017) show that this construction is one of the earliest emerging harbingers of complex syntax in Hebrew, as several phrases are conjoined to a single syntactic construct, containing or being contained by other syntactic devices. Headless relative clauses and relative clauses both attach to a nominal head and expand it while referring to the matrix nominal (Corrêa, 2018; Lau & Tanaka, 2021; O'Grady, 2011). Note how the next example demonstrates the Hebrew-specific interweaving of a relative clause with a conjunct structure: *there are children who tend to compare themselves to older children such as their siblings, relatives or neighbors.* The remaining construction that we focused on was complement clauses, subordinated clauses attached to a verb as in *several researchers claim that the use and accessibility of the cellular device affect the socialization habits of users.* Complement clauses are not difficult to produce or comprehend, and children produce them early on (Boeg Thomsen et al., 2021; Ögel-Balaban & Aksu-Koç 2020). Their importance lies in the type of predicates – especially mental verbs and adjectives – that govern such clauses (Brandt, 2020; Maekelberghe, 2021), promoting the ultimate development of Theory of Mind and social cognition.

In sum, the results of the current study underscore the critical importance of language knowledge and use, especially those that answer to the Typological Impact in serving the specific attributes of the language of young writers (Le Bruyn et al., 2022). Moreover, this section shows that complex syntax in its most mature form is found only in the adult texts. True, the constructions under study did appear in the children's and the adolescents' texts, but the only arena they congregated in large, statistically significant numbers were in the adults' texts, showing that the developmental processes we delineated during Later language acquisition culminate in maturity.

A genre perspective on the development of written text production

The overwhelming majority of our results were mitigated either by simple effects and / or interactions where genre – in this case expository and informative texts – was prominent. To begin with, given the absence of similar studies analyzing the linguistic features of these two particular genres in development, we did not have a particular direction on which to hang our hypotheses. The results have pinpointed the differences between the two genres in the acquisition of Hebrew writing skills. Table 9 summarizes the major genre- related differences detected in our study, showing in which genre each category type is more numerous.

Table 9: Summary of genre-dependent results of the current study.

Category	**Expository Texts**	**Informative Texts**
1. # Clauses	√	
2. Abstract noun tokens	√	
3. Abstract noun tokens in compounds	√	
4. Abstract noun types	√	
5. Abstract noun types in compounds	√	
6. Attributive adjective tokens		√
7. Attributive adjective types		√
8. Denominal attributive adjective tokens		√
9. Denominal attributive adjective types		√
10. Adjacency compound tokens		√
11. Chained compound tokens		√
12. Conjunct structures	√	
13. Complement clauses	√	
14. Headless clauses	√	
15. Relative clauses		√
16. Connective lexical markers		√
17. Commas		√

As the Results section and Table 9 show, the measures used in the current analysis were diagnostic in two perspectives – genre characteristics and the acquisition of genre-oriented writing. The linguistic categories we investigated had different distributions in the two genres, thus reflecting the overall character of either expositions or informative texts. Informational and expository texts are indeed close in what Berman (2005) designates as having academic and non-narrative *stance* - that is, being generic, non-personal, non-dynamic, and with no unfolding of events and no focal, overarching narrative point. However, the different distributions of the studied constructions indicate that these are two *different* non-narrative genres.

Expositions typically serve the discussion of social-related values, concepts and processes, and are highly abstract by nature. Their protagonists are ideas and propositions (Britton, 1994). In our case the expository theme we asked participants to write about was the notions of *success and failure*. It is no wonder that abstract nouns by any count as well as compounds based on abstract nouns were more numerous in expositions rather than in informative texts. The following array of abstract nouns occurred in one adult text (in addition to *success and failure*) – testifying to the abstract nature of the text: *musag* 'concept', *hevdel* 'difference', *matara* 'goal', *haga'a* 'attainment', *toca'a* 'result', *txum* 'domain', *histaklut* 'observation', *parshanut* 'interpretation', *meci'u*t 'reality', and *nisyonot* 'experiences'. As noted above, these abstract nouns function as attractors for complex syntactic constructions of various kinds, including compounding and noun-adjective structures such as *téva ha-adam* 'man's=human nature', *nifla'ot ha-maxshir* '(the) marvels (of) the-device'; *shalom pnimi* 'internal peace', *sviva enoshit* 'human environment'. This important characteristic of expositions underscores their abstract stance and tendency towards complex syntax. This characteristic is enhanced by the fact that clauses are longer in expositions than in informative texts, with all of the implications discussed above. Moreover, three out of four syntactic measures indicating syntactic complexity - conjunct constructions, complement clauses and headless relatives – were more numerous in expository texts, often working together to create discourse chunks based on complex syntactic combinations as in *it is necessary to find the suitable balance between the abilities of the child and the assignments required of him, so that he will experience feelings of success and sometimes also feelings of failure to some extent.*

A second set of linguistic features shed light on the characteristics of informative texts. Informative texts offer a specialized outlook on topics, objects, entities, personae, locations, phenomena and processes in the present and the past, providing objective information on their nature and properties (Giora, 1993). This makes informative texts highly relevant throughout the school years (van Rijk et al., 2017). Like expositions, informative texts take a generic outlook (even on the life history of specific people), and they are not driven by the psychological relationships and motivations, as are narratives (Ravid & Zilberbuch, 2003). In non-expert writers, informative texts typically start with an Aristotelian definition (e.g., *the cellular phone is a device that is used for communication in our modern times*) and then elaborate on its predicate, so that the whole of the text is such an elaboration (Watson, 1985). This essential character of informative texts calls for much description and elaboration in

[1] A word of caution here: The two genres are not always clearly separated in studies of the development of writing, and are sometimes used interchangeably, as in Ray and Meyer (2011).

static, generic stance, as is evident from two types of constructions that are more prevalent in informative texts – attributive adjectives and relative clauses. These two constructions elaborate upon a nominal nucleus, so that more information is provided about various aspects of the noun. For example, note the array of denominal attributive adjectives in *gormim politiyim umixariyim* '**political and commercial** factors', *informacya xiyunit* '**vital** information', *metavxim xiconiyim* '**external** mediators' – most often coupled with abstract nouns. And the following example of three relative clauses in a single long sentence shows to what extent the elaboration of the nominal nucleus adds relevant information: *a cellular telephone (mobile telephone) is a device* ***that is possible to carry on a person's body****, with a mechanism* ***that enables carrying a conversation between two people who are in possession of this device.*** These two constructions constitute two typical facets of the components of complex syntax that are found in informative texts and contribute to its information density (Ravid & Berman, 2006).

Note that here, too, fully mature informative texts in the sense of having all of the typical linguistic features delineated here were only found in the adult group, underscoring the culmination of Later language development. However, there was one property that is a harbinger and a companion of complex syntax that proved to be significantly more prevalent in the informative genre – demarcation markers in the form of lexical connectives such as *benosaf* 'in addition', *kmo xen* 'likewise', *lemashal* 'for example', *lefixax* 'therefore', as well as commas. These two features signify the organization of the text into coherent, logical and causal segments that contribute to the flow of information. These are two important properties of cohesive texts which tie together in various ways – elaborating, illustrating, comparing, delineating the pieces of information that make up the whole text. Clearly, young writers found it easier to produce more coherent, well demarcated informative texts than expositories, probably due to the more concrete nature of the former and the more difficult cognitive demands of the latter genre.

In conclusion, the study we have presented delved deep into the constituents of academic writing in a developmental perspective. We found that complex lexicon and syntax in discourse increase in prevalence across the school years, with specific structures being genre-typical as befits the stance and character of the two genres under investigation. We also found that it was only in adulthood that all of these components were employed in optimal fashion, at the culmination of the period of Later Language Development. Our results imply that informative and expository texts are indeed distinct in their characteristics, and that it takes many years of internal linguistic and cognitive development, on the one hand, and schooling instruction and experience, on the other hand, to achieve qualitative academic writing in non-experts.

References

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58*(10).

Aparici, M., Rosado, E., & Perera, J. (2016). Later development of relative clauses across discourse genres and modalities of production. In J. Perera, M. Aparici, E. Rosado, & N. Salas (Eds.), *Written and spoken language development across the lifespan* (pp. 201-225). Cham: Springer.

Ashkenazi, O. & Ravid, D. (1998). Children's understanding of linguistic humor: An aspect of metalinguistic awareness. *Current Psychology of Cognition, 17*, 367-387.

Atanasova, T., Fargier, R., Zesiger, P., & Laganaro, M. (2020). dynamics of word production in the transition from adolescence to adulthood. *Neurobiology of Language, 2*(1), 1-21.

Baaijen, V. M. & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction, 36*(3), 199-223.

Balthazar, A. H., & Scott, C. B. (2018). Targeting complex sentences in older school children with specific language impairment: results from an early phase treatment study. *Journal of Speech, Language, and Hearing Research, 61*, 713-728.

Bar-On, A. & Ravid, D. (2011). Morphological analysis in learning to read pseudowords in Hebrew. *Applied Psycholinguistics, 32*, 553-581.

Beers, F.B. & Nagy, W.E. (2011). Writing development in four genres from grades three to seven: syntactic complexity and genre differentiation. *Reading and Writing: An Interdisciplinary Journal, 24*, 183-202.

Beitchman, J.H., Jiang, H., Koyama, E., Johnson, C.J., Escobar, M., Atkinson, L., ... & Vida, R. (2008). Models and determinants of vocabulary growth from kindergarten to adulthood. *Journal of Child Psychology and Psychiatry, 49*, 626-634.

Beltrán-Planques, V., & Querol-Julián, M. (2018). English language learners' spoken interaction: What a multimodal perspective reveals about pragmatic competence. *System, 77*, 80-90.

Ben-Zvi, G. & Levie, R. (2016). Development of Hebrew derivational morphology from preschool to adolescence. In R.A. Berman (Ed.), *Acquisition and development of Hebrew: From infancy to Adolescence* (pp. 135-173). Amsterdam: Benjamins.

Berman, R.A. (2004). Between emergence and mastery: The long developmental route of language acquisition. In: *Language Development across Childhood and Adolescence.* (pp. 9-34). Amsterdam: John Benjamins.

Berman, R.A. (2005). Introduction: Developing discourse stance in different text types and languages. *Journal of Pragmatics, 37*, 105-124.

Berman, R.A. (2007). Developing language knowledge and language use across adolescence. In E. Hoff & M. Shatz (Eds.), *Handbook of Language Development* (pp. 346-367) London: Blackwell.

Berman , R.A. (2008). The psycholinguistics of developing text construction. *Journal of Child Language, 35*, 735-771.

Berman, R. A. (2009). Acquisition of compound constructions. In: R. Lieber & Stekauer, P. (Eds.), *Handbook of Compounding* (pp. 298-322). Oxford: Oxford University Press, .

Berman, R.A. (2014). Syntactic complexity: Development of clause-combining. In P. Brooks & V. Kempe, (eds.), *Encyclopedia of language development* (pp. 634-640). New York: Sage.

Berman, R.A. (2016). Linguistic literacy and later language development. In J. Perera, M. Aparici, E. Rosado, and N. Salas (Eds) *Written and spoken language development across* *the lifespan* (pp. 181-200). Springer, Cham,

Berman, R.A. (2017). Word class distinctiveness versus polycategoriality in Modern Hebrew: Psycholinguistic perspectives. In V. Vapnarsky & E. Veneziano (Eds.) *Lexical polycategoriality: Cross-linguistic, cross-theoretical, and language acquisition approaches* (pp. 343-376). Amsterdam: Benjamins.

Berman, R.A. (2018). Development of complex syntax: From early clause-combining to text-embedded syntactic packaging. In A. Bar-On & D. Ravid. *Handbook of Language Disorders* (pp. 235-256). Mouton de Gruyter.

Berman, R.A., & Katzenberger, I. (2004). Form and function in introducing narrative and expository texts: A developmental perspective. *Discourse Processes, 38*(1), 57-94.

Berman, R.A., & Nir-Sagiv B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes, 43* (2), 79-120.

Berman, R.A., & Nir-Sagiv B. (2009). The language of expository texts: Developmental perspectives. In M. Nippold & C. Scott (Eds.), *Expository discourse in children, adolescents, and adults: Development and disorders* (pp. 101-123). NY: Taylor & Francis.

Berman, R.A. & Nir-Sagiv B. (2010) The lexicon in speech-writing differentiation: Developmental perspectives. *Written Language & Literacy, 13*, 181-203.

Berman, R.A., R. Nayditz, & D. Ravid. (2011). Linguistic diagnostics of written texts in two school-age populations. *Written Language & Literacy, 14*, 161-187.

Berman, R.A. & Ravid, D. (2008). Becoming a literate language user: Oral and written text construction across adolescence. In D. R. Olson and N. Torrance (Eds.), *Cambridge handbook of literacy* (pp. 92-111). Cambridge: Cambridge University Press.

Berman, R.A. & Ravid, D. (2010). Interpretation and recall of proverbs in three pre-adolescent populations. *First Language, 30*, 155-173.

Berninger, V., Abbott, R., Cook, C. R., & Nagy, W. (2017). Relationships of attention and executive functions to oral language, reading, and writing skills and systems in middle childhood and early adolescence. *Journal of Learning Disabilities, 50*(4), 434-449.

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*(2), 137-164.

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics, 37*(5), 639-668.

Blakemore, S. J. (2012). Imaging brain development: the adolescent brain. *Neuroimage, 61*(2), 397-406.

Boeg Thomsen, D., Theakston, A., Kandemirci, B., & Brandt, S. (2021). Do complement clauses really support false-belief reasoning? A longitudinal study with English-speaking 2-to 3-year-olds. *Developmental psychology, 57*(8), 1210.

Bolinger, D. (1967). Adjectives in English: attribution and predication. *Lingua, 18*, 1-34.

Brandes, G. & Ravid, D. (2016). Prepositional phrases as manner adverbials in the development of Hebrew L1 text production. In: Ortega L., Tyler A. E., Park, H. I. & Uno M. (Eds.), *The usage-based study of language learning and multilingualism* (pp. 55-73). Washington D.C.: Georgetown University Press.

Brandes, G. & Ravid, D. (2019). The development of adverbial clause functions in Hebrew narrative and expository writing across adolescence. *Written Language and Literacy, 22*(1), 130-158.

Brandt, S. (2020). Social cognitive and later language acquisition. *Current perspectives in child language acquisition: how children use their environment to learn,* 155-170.

Brimo, D., & Hall-Mills, S. (2019). Adolescents' production of complex syntax in spoken and written expository and persuasive genres. *Clinical Linguistics & Phonetics, 33*(3), 237-255.

Britton, B. K. (1994). Understanding expository text: Building mental structures to induce insights. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641-674). Academic Press.

Bybee, J. L., & Noonan, M. (Eds.). (2002). *Complex sentences in grammar and discourse: Essays in honor of Sandra A. Thompson.* Amsterdam: Benjamins.

Cahana-Amitay, D., & Ravid, D. (2000). Optional bound morphology in the development of text production. In S.C. Howell, S.A. Fish & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (Vol. I, pp. 176-184). Somerville, MA: Cascadilla Press.

Castillo, C. & Tolchinsky, L. (2018). The contribution of vocabulary knowledge and semantic orthographic fluency to text quality through elementary school in Catalan. *Reading and Writing, 31*(2), 293-323.

Chafe, W. L. (1994). *Discourse, consciousness, and time: the flow of language in speech and writing.* Chicago: Chicago University Press.

Clark, E. V. (2004). How language acquisition builds on cognitive development. *Trends in Cognitive Sciences, 8*(10), 472-478.

Clark, E. V. (2017). Later lexical development and word formation. In P. Fletcher and B. MacWhinney (Eds.), *The handbook of child language* (pp. 393-412). Oxford: Blackwell.

Colston, H.L. (2020). Figurative language development/acquisition research: Status and ways forward. *Journal of Pragmatics, 156,* 176-190.

Corrêa, L.M.S. (2018). The relative difficulty of children's comprehension of relative clauses: A procedural account. In *Children's language* (pp. 225-243). Psychology Press.

Crone, E.A. (2009). Executive functions in adolescence: inferences from brain and Behavior. *Developmental Science, 12,* 825-830.

Crossley, S.A., Weston, J.L., McLain Sullivan, S.T. & McNamara, D.S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28,* 282-311.

Cutillas, L., & Tolchinsky, L. (2017). Use of adjectives in Catalan: A morphological characterization in different genres and modes of production through school-age development. *First Language, 37*(1), 58-82.

Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C., & Bivens-Tatum, J. (2008). *Cognitive models of writing: Writing proficiency as a complex integrated skill.* ETS Research Report Series.

Delage, H., Stanford, E., & Durrleman, S. (2021). Working memory training enhances complex syntax in children with Developmental Language Disorder. *Applied Psycholin-*

guistics, 42(5), 1341-1375.
De Ruiter, L., Theakston, A., Brandt, S., & Lieven, E. (2018). Iconicity affects children's comprehension of complex sentences. *Cognition, 171,* 202-224.
Donovan, C.A. (2001). Children's development and control of written story and informational genres: Insights from one elementary school. *Research in the Teaching of English,* 394-447.
Donovan, C. A., & Smolkin, L. B. (2001). Genre and other factors influencing teachers' book selections for science instruction. *Reading Research Quarterly, 36*(4), 412-440.
Egbert, J. (2020). The new normal?: A pandemic of task engagement in language learning. *Foreign Language Annals, 53*(2), 314-319.
Elsabbagh, M. & Karmiloff-Smith, A. (2004). Modularity of mind and language. In Brown, K. (ed.), *Encyclopedia of language and linguistics* (pp. 218-222). Amsterdam: Elsevier.
Ferris, C. (2014). *The meaning of syntax: A study in the adjectives of English.* Routledge.
Figueroa, J., Meneses, A., & Chandia, E. (2018). Academic language and the quality of written arguments and explanations of Chilean 8th graders. *Reading and Writing, 31*(3), 703-723.
Fortman, J. (2003). Adolescent language and communication from an intergroup perspective. *Journal of Language and Social Psychology 22*(1), 104-111.
Fuhrmann, D., Knoll, L. J., & Blakemore, S. J. (2015). Adolescence as a sensitive period of brain development. *Trends in Cognitive Sciences, 19*(10), 558-566.
Gillam, R. B., & Johnston, J. R. (1992). Spoken and written language relationships in language / learning-impaired and normally achieving school-age children. *Journal of Speech, Language, and Hearing Research, 35*(6), 1303-1315.
Giora, R. (1993). On the function of analogies in informative texts. *Discourse Processes, 16*(4), 591-611.
Graham, S. & Harris, K.H. (2009). Almost 30 years of writing research: making sense of it all with the wrath of khan. *Learning Disabilities Research & Practice, 24,* 58-68.
Gregg, N. & Nelson, J. (2018). Empirical Studies on theWriting Abilities of Adolescents and Adults with Learning Difficulties. In: Miller, .B, McCardle, P. & Connelly, V. (Eds.), *Approaches to Improving Writing Research, Instruction, and Performance. Understanding the Needs of Writers across the Life course* (pp. 73-98). Brill.
Hayes, J. R., & Flower, L. (1981). *Uncovering cognitive processes in writing: An introduction to protocol analysis.* ERIC Clearinghouse.
Jisa, H. & Tolchinsky, L. (2009). Developing a depersonalized stance through linguistic means in typologically different languages: Written expository discourse. *Written Language & Literacy, 12*(1), 1-25.
Kadosh, K. C., Linden, D. E., & Lau, J. Y. (2013). Plasticity during childhood and adolescence: innovative approaches to investigating neurocognitive development. *Developmental Science, 16*(4), 574-583.
Kärkäinen, E., Sorjonen, M. L., & Helasvuo, M. L. (2007). Discourse structure. In *Language typology and syntactic description: Volume II: Complex constructions* (pp. 301-371). Cambridge: Cambridge University Press.
Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive*

science. Cambridge: MIT Press.

Kawar, K. (2021). Morphology and Syntax in Arabic-Speaking Adolescents Who Are Deaf and Hard of Hearing. *Journal of Speech, Language, and Hearing Research, 64*(10), 3867-3882.

Kellogg, R.T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research, 1*, 1-26.

Khokhlova, N. (2014). Understanding of abstract nouns in linguistic disciplines. *Procedia-Social and Behavioral Sciences, 136*, 8-11.

Kilford, E. J., Garrett, E., & Blakemore, S. J. (2016). The development of social cognition in adolescence: An integrated perspective. *Neuroscience & Biobehavioral Reviews, 70*, 106-120.

Kuhn, D. (2011). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Handbook of childhood cognitive development* (2nd ed., pp. 472-523). Oxford: Blackwell

Larsen, B., & Luna, B. (2018). Adolescence as a neurobiological critical period for the development of higher-order cognition. *Neuroscience & Biobehavioral Reviews, 94*, 179-195.

Lau, E., & Tanaka, N. (2021). The subject advantage in relative clauses: A review. *Glossa, 6*(1), 1-34.

Le Bruyn, B., Fuchs, M., van der Klis, M., Liu, J., Mo, C., Tellings, J., & De Swart, H. (2022). Parallel corpus research and target language representativeness: The contrastive, typological, and translation mining traditions. *Languages, 7*(3), 176.

Lecce, S., Caputi, M., Pagnin, A., & Banerjee, R. (2017). Theory of mind and school achievement: The mediating role of social competence. *Cognitive Development, 44*, 85-97.

Lindgren, E., Leijten, M., & Van Waes, L. (2011). Adapting to the reader during writing. *Written Language & Literacy, 14*(2), 188-223.

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 26, No. 26).

Maekelberghe, C. (2021). Assessing the impact of syntactic complement type on the modifying status of complement-taking predicates: the case of imagine. *Language Sciences, 86*, 101390

McNamara, D.S., Crossley, S.A. & McCarthy, P.M. (2010). Linguistic features of writing quality. *Written Communication, 27*, 57-86.

Murray, N. (2012). *Writing essays in English language and linguistics: Principles, tips and strategies for undergraduates.* Cambridge: Cambridge University Press.

Nelson, K. (1976). Some attributes of adjectives used by young children. *Cognition, 4*(1), 13-30.

Nelson, N.W. (2013). Syntax development in the school-age years: Implications for assessment and intervention. *Perspectives on Language and Literacy, 39*(3), 9-17.

Nippold, M.A. (2016). *Later language development: School-age children, adolescents, and young adults.* Austin, TX: PRO-ED.

Nippold, M.A., Hegel, S. L., Sohlberg, M. M., & Schwarz, I. E. (1999). Defining abstract entities: Development in pre-adolescents, adolescents, and young adults. *Journal of*

Speech, Language, and Hearing Research, 42(2), 473-481.

Nippold, M.A., Mansfield, T.C., Billow, J.S. & Tomblin, J.B. (2008). Expository discourse in adolescents with language impairments: Examining Syntactic Development. *American Journal of Speech-Language Pathology, 17*, 356-366.

Nippold, M.A., Mansfield, T.C., Billow, J.S. & Tomblin, J.B. (2009). Syntactic development in adolescents with a history of language impairments: A follow-up investigation. *American Journal of Speech-Language Pathology, 18*, 241-251.

Nippold, M.A., & Scott, C.M. (Eds.) (2010). *Expository discourse in children, adolescents, and adults: Development and disorders.* New York, NY: Psychology Press.

Nippold, M. A., & Sun, L. (2010). Expository writing in children and adolescents: A classroom assessment tool. *Perspectives on Language Learning and Education, 17*(3), 100-107.

Nippold, M.A, Ward-Lonergan, J.M. & Fanning, J.L. (2005). Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic, and pragmatic development. *Language Speech and Hearing Services in Schools, 36*, 125-138.

Northey, M., McCutchen, D., & Sanders, E. A. (2016). Contributions of morphological skill to children's essay writing. *Reading and writing, 29*(1), 47-68.

Ögel-Balaban, H., & Aksu-Koç, A. (2020). Clause Chaining and Discourse Continuity in Turkish Children's Narratives. *Frontiers in psychology, 11*, 115.

O'Grady, W. (2011). Relative clauses: Processing and acquisition. *The acquisition of relative clauses: Processing, typology and function*, 13-38.

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing, 26*, 45-65.

Olson, D.R. (1994). *The world on paper: The conceptual and cognitive implication of writing and reading.* Cambridge: Cambridge University Press.

Olson, D.R. (2006). Oral discourse in a world of literacy. *Research in the Teaching of English, 41*(2), 136-143.

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition, 126*(2), 268-284.

Osterhaus, C., Koerber. S., & Sodian, B. (2017). Scientific thinking in elementary school: Children's social cognition and their epistemological understanding promote experimentation skills. *Developmental Psychology, 53*, 450-62.

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes, 14*, 48-59.

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences, 9*(2), 60-68.

Prados, M. D. (2018). Abstract nouns as metadiscursive shells in academic discourse. Caplletra. *Revista Internacional de Filologia, 64*, 153-178.

Proverbio, A.M., & Zani, A. (2005). Developmental changes in the linguistic brain after puberty. *Trends in Cognitive Sciences, 9*, 164-167.

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences, 11*(7), 274-279.

Ramscar, M., Dye, M., Blevins, J., & Baayen, H. (2018). Morphological development. In A. Bar-On and D. Ravid (Eds.), *Handbook of Communications Disorders: Theoretical, Empirical, and Applied Linguistic Perspectives* (pp. 181-202). Berlin: de Gruyter.

Ravid, D. (2004). Later lexical development in Hebrew: derivational morphology revisited. In R.A. Berman (Ed.) *Language development across childhood and adolescence: Psycholinguistic and crosslinguistic perspectives* (pp. 53-82). Amsterdam: Benjamins.

Ravid, D. (2005). Emergence of linguistic complexity in written expository texts: Evidence from later language acquisition. In: Ravid, D. & Bat-Zeev Shyldkrot, H. (Eds.), *Perspectives on language and language development* (pp. 337-355). Dordrecht: Kluwer.

Ravid, D. (2006). Semantic development in textual contexts during the school years: Noun scale analyses. *Journal of Child Language, 33*, 791-821.

Ravid, D. (2013). Syntactic complexity in discourse production across different text types. In: Bolly, C. & Degand, L. (Eds.), *Text-structuring. Across the line of speech and writing variation* (Corpora and Language in Use series No. 2). Louvain-la-Neuve: Presses Universitaires de Louvain.

Ravid, D. (2019). Derivation. In R. A. Berman (Ed.), *Usage-based studies in Modern Hebrew: Background, morpho-lexicon, and syntax* (pp. 203-265). Amsterdam: Benjamins.

Ravid, D. & Assulin Tzabar, N. (2017). Compounding in early child speech: Hebrew peer talk 2-8. In Wolfgang U. Dressler, F. Nihan Ketrez and Marianne Kilani Schoch (Eds.), *Nominal compound acquisition* (pp. 251-274). Amsterdam: Benjamins.

Ravid, D. & Avidor, A. (1998). Acquisition of derived nominals in Hebrew: developmental and linguistic principles. *Journal of Child Language, 25*, 229-266.

Ravid, D., Bar-On, A., Levie, R. & Douani, O. (2016). Hebrew adjective lexicons in developmental perspective: Subjective register and morphology. *The Mental Lexicon, 11*, 401-428.

Ravid, D. &. Berman, R.A. (2006). Information density in the development of spoken and written narratives in English and Hebrew. *Discourse Processes, 41*, 117-149.

Ravid, D. &. Berman, R.A. (2009). Developing linguistic register across text types: The case of Modern Hebrew. *Pragmatics and Cognition, 17*, pp. 108-145.

Ravid, D. & Berman, R.A. (2010). Developing noun phrase complexity at school age: a text-embedded cross-linguistic analysis. *First Language, 30*, 3-26.

Ravid, D. & Cahana-Amitay, D. (2005). Verbal and nominal expression in narrating conflict situations in Hebrew. *Journal of Pragmatics, 37*, 157-183.

Ravid, D., Dromi, E. & Kotler, P. (2009). Linguistic complexity in school-age text production: Expository vs. mathematical discourse. In M. A. Nippold & C. M. Scott, (Eds.), *Expository discourse in children, adolescents, and adults: Development and disorders* (pp. 125-156). NY: Psychology Press.

Ravid, D. & Hershkovitz, L. (2017). The development of Hebrew conjunct constructions in narration. In Eliane Segers and Paul van den Broek (Eds.) *Developmental perspectives in written language and literacy. In honor of Ludo Verhoeven* (pp. 119-135). Amsterdam: Benjamins.

Ravid, D. & Levie, R. (2010). Adjectives in the development of text production: Lexical, morphological and syntactic analyses. *First Language, 30*, 27-55.

Ravid, D. & Shalom, T. (2012). *The Standards report. Final report to Yad Hanadiv* (Rothschild Fund). Tel Aviv: CET and Tel Aviv University. [in Hebrew]

Ravid, D. & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language, 29*, 419-448.

Ravid, D., van Hell, J., Rosado, E., & Zamora, A. (2002). Subject NP patterning in the development of text production: Speech and writing. *Written Language and Literacy, 5*, 69-94.

Ravid, D. & S. Zilberbuch. (2003). The development of complex nominals in expert and non-expert writing: A comparative study. *Pragmatics and Cognition, 11*(2), 267-297.

Ray, M. N., & Meyer, B. J. (2011). Individual differences in children's knowledge of expository text structures: A review of literature. *International Electronic Journal of Elementary Education, 4*(1), 67-82.

Salas, N., & Caravolas, M. (2019). Dimensionality of Early Writing in English and Spanish. *Journal of Literacy Research, 51*(3), 272-292.

Sanders, T.J.M. & Schiperoord, J. (2006). Text structure as a window on the cognition of writing: How text analysis provides insights in writing products and writing processes. In: MacArthur, C.H., Graham, S. & Fitzgerald, J. (Eds.), *Handbook of Writing Research* (pp. 386-402). NY: The Guilford Press.

Schmid, H. J. (2012). English abstract nouns as conceptual shells. In *English Abstract nouns as conceptual shells.* De Gruyter Mouton.

Schleppegrell, M.J. (2001). Linguistic features of the language of schooling. *Linguistics and Education, 12*, 431-459.

Schuele, C.M., & Dykes, J.C. (2005). Complex syntax acquisition: a longitudinal case study of a child with specific language impairment. *Clinical Linguistics & Phonetics, 19*(4), 295-318.

Scott, C.M. & Balthazar, C.H. (2010). The grammar of information: Challenges for older students with language impairments. *Topics in Language Disorders, 30*, 288-307.

Scott, C.M. & Windsor, J. (2000). Measures in spoken and written narrative and expository discourse of school aged children with language learning disabilities. *Journal of Speech Language and Hearing Research, 43*, 324-339.

Silliman, E. R., & Berninger, V. W. (2011). Cross-disciplinary dialogue about the nature of oral and written language problems in the context of developmental, academic, and phenotypic profiles. *Topics in Language Disorders, 31*(1), 6-23.

Simmonds, D.J., Hallquist, M.N., Asato, M., Luna, B., (2014). Developmental stages and sex differences of white matter and behavioral development through adolescence: a longitudinal diffusion tensor imaging (DTI) study. *NeuroImage, 92*, 356-368.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. *The Cambridge handbook of literacy* (pp. 112 - 133).

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication, 33*(2), 149-183.

Tibi, S., Tock, J. L., & Kirby, J. R. (2019). The development of a measure of root awareness to account for reading performance in the Arabic language: A development and valida-

tion study. *Applied Psycholinguistics, 40*(2), 303-322.
Tolchinsky, L. (2019). Evolving structure of descriptive texts and learners' abilities. *Journal of Literacy Research, 51*(3), 293-314.
Tower, C. (2003). Genre development and elementary students' informational writing: A review of the literature. *Literacy Research and Instruction, 42*(4), 14-39.
van Rijk, Y., de Mey, L., de Haan, D., van Oers, B., & Volman, M. (2017). Reading for meaning: the effects of developmental education on motivation and achievement in reading informative texts in primary school. *ReseaRch PaPeRs in education, 32*(3), 333-352.
van Rijt, J. H., van den Broek, B., & De Maeyer, S. (2021). Syntactic predictors for text quality in Dutch upper-secondary school students' L1 argumentative writing. *Reading and Writing, 34*(2), 449-465.
Vulchanova, M., Milburn, E., Vulchanov, V., & Baggio, G. (2019). Boon or burden? The role of compositional meaning in figurative language processing and acquisition. *Journal of Logic, Language and Information, 28*(2), 359-387.
Watson, R. (1985). Towards a theory of definition. *Journal of child language, 12*(1), 181-197.

Appendix A. GEE models of text size

(i) Number of Words – GEE Modeling Results, Main Effects and Interactions

Model Type: NBD Outcome Variable: Number of Words		
Step 1: Main Effects	Exp(B)	Coefficient
Intercept	66.77	4.20*** (.05)
Grade Effect		**Wald χ^2 = 295.2, *p*<.001**
G4 vs. Adults	.51	-.68*** (.08)
G7 vs. Adults	.75	-.29*** (.07)
G11 vs. Adults	1.65	.50*** (.07)
Genre Effect		**Wald χ^2 = .72, *p*=.40**
Informative vs. Narrative	1.02	.02 (.02)
Step 2: Interaction		Wald χ^2= 12.13, *p*=.007

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

(ii) Number of Clauses – GEE Modeling Results, Main Effects and Interaction

Model Type: NBD Outcome Variable: Number of Clauses		
Step 1: Main Effects	Exp(B)	Coefficient
Intercept	12.57	2.53*** (.05)
Grade Effect		**Wald χ^2 = 191.20, *p*<.01**
G4 vs. Adults	.78	-.25*** (.07)
G7 vs. Adults	.88	-.01*** (.07)
G11 vs. Adults	1.89	.64*** (.07)
Genre Effect		**Wald χ^2 = 48.35, *p*=<.001**
	0.83	-.19 (.03)
Step 2: Interaction		Wald χ^2= 12.13, *p*=.007

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

(iii) Mean Clause Length – GEE Modeling Results, Main Effects and Interaction

Model Type: Gamma Outcome Variable: Mean Clause Length	
Step 1: Main Effects	Coefficient
Intercept	5.57*** (.12)
Grade Effect	**Wald χ^2 = 192.99, *p*<.0001**
G4 vs. Adults	-2.01*** (.15)
G7 vs. Adults	-1.20*** (.18)
G11 vs. Adults	-.77*** (.13)
Genre Effect	**Wald χ^2 = 71.52, *p*=<.001**
	.86*** (.10)
Step 2: Interaction	Wald χ^2= 20.45, *p*<.001

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Appendix B. GEE models of Lexical measures

(i) Abstract noun tokens over Words– GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Abstract noun tokens over Words	
Step 1: Main Effects	Coefficient
Intercept	-1.48*** (.083)
Grade Effect	**Wald χ^2 = 57.83, *p*<.001**
G4 vs. Adults	-.30*** (.05)
G7 vs. Adults	-.26*** (.04)
G11 vs. Adults	-.15** (.32)
Genre Effect	**Wald χ^2 = 271.05, *p*<.001**
Informative vs. Narrative	-.67*** (.001)
Step 2: Interaction	Wald χ^2= 7.71, *p*=.052

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Abstract noun tokens in compounds over words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Abstract noun tokens in compounds over words	
Step 1: Main Effects	Coefficient
Intercept	-2.52*** (.051)
Grade Effect	**Wald χ^2 = 252.752, *p*<.001**
G4 vs. Adults	-.30*** (.11)
G7 vs. Adults	-.26*** (.09)
G11 vs. Adults	-.15** (.06)
Genre Effect	**Wald χ^2 = 46.56, *p*<.001**
Informative vs. Narrative	.19*** (.0013)
Step 2: Interaction	Wald χ^2= 2.49, *p*=.478

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Abstract noun types over Words – GEE Modeling Results, Main Effects and InteractionsAbstract noun types over Words– GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Abstract noun types over Words	
Step 1: Main Effects	Coefficient
Intercept	-1.80*** (.03)
Grade Effect	**Wald $\chi^2 = 65.11$, *p*<.001**
G4 vs. Adults	-.30*** (.05)
G7 vs. Adults	-.26*** (.05)
G11 vs. Adults	-.26** (.04)
Genre Effect	**Wald $\chi^2 = 37.15$, *p*<.001**
Informative vs. Narrative	-.21*** (.003)
Step 2: Interaction	Wald χ^2= 25.09, *p*<.001

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Abstract noun types in compounds over words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Abstract noun types in compounds over words	
Step 1: Main Effects	Coefficient
Intercept	-1.99*** (.03)
Grade Effect	**Wald $\chi^2 = 284.77$, *p*<.001**
G4 vs. Adults	-1.52*** (.11)
G7 vs. Adults	-1.08*** (.09)
G11 vs. Adults	-.45*** (.07)
Genre Effect	**Wald $\chi^2 = 54.17$, *p*<.001**
Informative vs. Narrative	-.027*** (.000)
Step 2: Interaction	Wald χ^2= 17.97, *p*<.001

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

(ii) Attributive adjective tokens over Words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Attributive adjective tokens over Words	
Step 1: Main Effects	Coefficient
Intercept	-3.20*** (.05)
Grade Effect	**Wald χ^2 = 35.09, *p*<.001**
G4 vs. Adults	-.39*** (.08)
G7 vs. Adults	-.35*** (.07)
G11 vs. Adults	-.18** (.06)
Genre Effect	**Wald χ^2 = 47.89, *p*<.001**
Informative vs. Narrative	.013*** (.002)
Step 2: Interaction	Wald χ^2= 5.66, *p*=.13

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Attributive adjective types over Words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Attributive adjective tokens over Words	
Step 1: Main Effects	Coefficient
Intercept	-3.31*** (.05)
Grade Effect	**Wald χ^2 = 45.07, *p*<.001**
G4 vs. Adults	-.43*** (.08)
G7 vs. Adults	-.38*** (.07)
G11 vs. Adults	-.22*** (.05)
Genre Effect	**Wald χ^2 = 49.78, *p*<.001**
Informative vs. Narrative	.011*** (.002)
Step 2: Interaction	Wald χ^2= 1.41, *p*=.70

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Denominal attributive adjective tokens over Words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Denominal attributive adjective tokens over Words	
Step 1: Main Effects	Coefficient
Intercept	-4.06*** (.08)
Grade Effect	**Wald χ^2 = 232.18, *p*<.001**
G4 vs. Adults	-1.77*** (.15)
G7 vs. Adults	-1.09*** (.09)
G11 vs. Adults	-.54** (.08)
Genre Effect	**Wald χ^2 = 366.32, *p*<.001**
Informative vs. Narrative	.022*** (.001)
Step 2: Interaction	Wald χ^2= 5.66, *p*=.13

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Denominal attributive adjective types over Words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Denominal attributive adjective types over Words	
Step 1: Main Effects	Coefficient
Intercept	-4.06*** (.09)
Grade Effect	**Wald χ^2 = 271.09, *p*<.001**
G4 vs. Adults	-1.78*** (.15)
G7 vs. Adults	-1.16*** (.09)
G11 vs. Adults	-.70*** (.08)
Genre Effect	**Wald χ^2 = 242.92, *p*<.001**
Informative vs. Narrative	.014*** (.001)
Step 2: Interaction	Wald χ^2= 26.82, *p*<.001

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

(iii) Adjacency compound tokens over Words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Adjacency compound tokens over Words	
Step 1: Main Effects	Coefficient
Intercept	-3.15*** (.05)
Grade Effect	**Wald χ^2 = 133.06, *p*<.001**
G4 vs. Adults	-.99*** (.10)
G7 vs. Adults	-.64*** (.10)
G11 vs. Adults	-.18** (.06)
Genre Effect	**Wald χ^2 = 97.12, *p*<.001**
Informative vs. Narrative	.19*** (.002)
Step 2: Interaction	Wald χ^2= 5.17, *p*=.16

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Free compound tokens over Words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Free compound tokens over words	
Step 1: Main Effects	Coefficient
Intercept	-4.62*** (.1)
Grade Effect	**Wald χ^2 = 29.84, *p*<.001**
G4 vs. Adults	-.79*** (.15)
G7 vs. Adults	-.54*** (.10)
G11 vs. Adults	-.03 (.13)
Genre Effect	**Wald χ^2 = .17, *p*<.001**
Informative vs. Narrative	.00 (.00)
Step 2: Interaction	Wald χ^2= 9.9, *p*<.05

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Chained compound tokens over Words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Chained compound tokens over Words	
Step 1: Main Effects	Coefficient
Intercept	-4.83*** (.17)
Grade Effect	**Wald χ^2 = 47.33, *p*<.001**
G4 vs. Adults	-1.34*** (.28)
G7 vs. Adults	-1.36*** (.27)
G11 vs. Adults	-.62*** (.15)
Genre Effect	**Wald χ^2 = 32.52, *p*<.001**
Informative vs. Narrative	.005*** (.001)
Step 2: Interaction	Wald χ^2= 5.10, *p*=.17

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Appendix C. GEE models of Demarcation variables

(i) Conjoining lexicon over Words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Conjoining lexicon tokens over Words	
Step 1: Main Effects	Coefficient
Intercept	-3.73*** (.03)
Grade Effect	**Wald χ^2 = 19.94, *p*<.001**
G4 vs. Adults	.20* (.10)
G7 vs. Adults	.27** (.09)
G11 vs. Adults	-.08 (.08)
Genre Effect	**Wald χ^2 = 22.76, *p*<.001**
Informative vs. Narrative	.01*** (.00)
Step 2: Interaction	Wald χ^2= 2.60, *p*=.46

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Commas over words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear Outcome Variable: Commas over Words	
Step 1: Main Effects	Coefficient
Intercept	-2.90*** (.09)
Grade Effect	**Wald χ^2 = 50.18, *p*<.001**
G4 vs. Adults	-1.23*** (.10)
G7 vs. Adults	-.26* (.12)
G11 vs. Adults	-.04 (.09)
Genre Effect	**Wald χ^2 = 5.84, *p*<.05**
Informative vs. Narrative	.006** (.00)
Step 2: Interaction	Wald χ^2= 2.94, *p*=.40

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

Full stops over Words – GEE Modeling Results, Main Effects and Interactions

Model Type: Transformed Linear	
Outcome Variable: Full stops over Words	
Step 1: Main Effects	Coefficient
Intercept	-2.75*** (.03)
Grade Effect	**Wald χ^2 = 19.15, *p*<.001**
G4 vs. Adults	-.33*** (.08)
G7 vs. Adults	-.16* (.06)
G11 vs. Adults	-.06 (.05)
Genre Effect	**Wald χ^2 = .28, *p*=.59**
Informative vs. Narrative	.006** (.00)
Step 2: Interaction	Wald χ^2= 6.37, *p*=.10

*** *p*<.001, ** *p*<.01, * *p*<.05; Standard errors in parentheses.

STEM-, SPRAAK- EN TAALPATHOLOGIE
Vol. 27, 2022, Special Issue, pp. 135-144
Special Issue for Steven Gillis
Gepubliceerd online: 5 november 2022

University of Groningen Press
https://doi.org/10.21827/32.8310/2022-SG-135

Remarks on the medical and social models of research in deafness and language development

Gary Morgan

Universitat Oberta de Catalunya, Psychology and Education Sciences, Barcelona, Spain

Samenvatting

Onderzoek naar de taalontwikkeling van dove kinderen heeft een lange en complexe geschiedenis. Het werk op dit terrein wordt gemotiveerd door schijnbaar onverenigbare modellen voor wat doofheid betekent. Aan de ene kant documenteert het dominante medische model gehoorverlies en tekortkomingen in de gesproken taal. Zulk onderzoek draagt bij aan voortdurende verbeteringen in de gesproken taal als resultaat van neonatale screening en vroeg aangebrachte cochleaire implantaten. Aan de andere kant hebben de onderzoekers (geringer in aantal) die kijken naar doofheid en taalontwikkeling in het sociale model gepleit voor de diversiteit van dove kinderen, voor hun recht om gebarentalen te leren en te worden onderwezen in tweetalige scholen. Dit artikel bestrijkt een selectie van onderzoeksstudies over doofheid en taalontwikkeling vanuit zowel de medische als sociale modellen. Het belangrijkste doel van het artikel is om enkele opmerkingen te maken over een aantal standpunten die door onderzoekers zijn ingenomen en die om een meer zorgvuldige bespreking vragen, zodat ze het veld verder kunnen helpen. Het besluit met een suggestie voor hoe de twee divergerende modellen meer zouden kunnen samenkomen. Het voorstel is om de aandacht te vestigen op de factoren die leiden tot communicatieve interacties van een hoge kwaliteit in plaats van louter toegang tot woorden of gebaren.

Abstract

Research on deaf children's language development has a long and complex history. Work is motivated by seemingly incompatible models of what deafness means. On the one hand, the dominant medical model documents hearing loss and spoken language deficits. Research contributes to continuing improvements in spoken language outcomes following neo-natal screening and early cochlear implants. On the other hand, the smaller number of researchers looking at deafness and language development in the social model have championed the diversity of deaf children, their rights to learn signed languages and be educated in bilingual schools. This paper covers a selection of research studies

Correspondentieadres:
Gary Morgan
Universitat Oberta de Catalunya
Psychology and Education Sciences
Barcelona, Spain
E-mail: gmorgan@uoc.edu

on deafness and language development coming from both the medical and social models. The main objective of the paper is to offer some remarks concerning a set of standpoints taken by researchers which require more careful discussion in order to further the field. It concludes with a suggestion for how the two diverging models could converge more. The proposal is to focus attention on the factors which lead to high quality early communicative interactions rather than access to words or signs.

Preface

Steven Gillis has published widely on child language development including the acquisition of Dutch, phonological and morphological development (e.g. Van Severen, et al. 2013), theories of acquisition (e.g. Cassani, Grimm, Daelemans & Gillis, 2018), the role of the input (e.g. Odijk & Gillis, 2022) and in terms of the current paper many studies of childhood deafness and language development (e.g. Grandon, B. Vilain, Gillis, 2019; Boonen, Kloots, Nurzia & Gillis, 2021). His deafness research includes studies of parent child interaction, phonological, grammar and vocabulary development. His research is grounded in mainstream language development theory, in particular phonological aspects of language and his development of coding and analysis systems including the Child language data exchange system (CHILDES). This grounding of deafness in wider arguments concerning language learning is very fruitful for both the small deafness field and the wider cross-linguistic comparisons of language acquisition in different contexts. His approach is to compare the development of the building blocks of speech and language in hearing and deaf children and propose explanations for why there are similarities and differences.

Childhood deafness and language development

In the United Kingdom and many other western countries, 1-2 in 1000 children are born deaf (NICE, 2019). Congenital deafness therefore is a very low incidence developmental condition compared to, for example, Developmental Language Disorder (DLD, Bishop 2007) which affects 7-10% of children. Despite its low incidence rate, deafness has serious impacts on children's communication development and ensuing psycho-social milestones (Coene, Schauwers, Gillis, Rooryck & Govaerts, 2011; Theunissen, et al, 2014; Hoffman, Cejas & Quittner, 2016). Because of the implementation of neo-natal screening today, most congenital deafness is identified in the first weeks of life with families generally entering medical intervention programmes early and typically with hearing aids and then cochlear implants (CIs). An early intervention means that infants can begin to adapt to hearing the world (Levine, Strother-Garcia, Hirsh-Pasek, & Golinkoff, 2016). This has not always been the case, in the past, many children were not diagnosed deaf until significantly later and used hearing aids which provided lower quality access to sound and parental spoken communication. Consequently, it was common for Deaf/Hard of Hearing (DHH) children to have very delayed language development (Marschark & Spencer, 2006). In contrast, in current research and clinical practice, the majority of DHH children have age-appropriate spoken language development compared with their hearing peers (Dettman et al, 2016). Despite this progress some DHH children's language development continues to be variable. It is estimated that around 30% of children still experience delays (Bruijnzeel, Ziylan, Stegeman, Topsakal & Grolman, 2016).

Addressing the root cause of this large variability is the key topic of this paper. There is a debate as to why DHH children continue to have delays in language development between two broad groups of researchers in the medical and social models (e.g. Geers, et al. 2017 vs Hall, Hall & Caselli, 2019). These two models are often at loggerheads with very little cross-communication.

Medical versus social models of research on deafness and language development

In the wider literature two main models are prominent in understanding disability: the medical model and the social model (Marks, 1997). Briefly, the medical model states that disability is a 'problem' within the person, caused by disease, trauma, or other health conditions and therefore requires sustained medical care and a cure. Whereas the medical model sees disability as a problem with the person, the social model considers the problem within wider society. The social model sees "disability" as a socially constructed problem in the environment that prevents disabled people from fully integrating. In the social model, disability is not an attribute of the individual but rather a complex collection of conditions created by the social environment. These two models applied to the study of deafness mean very different things for researchers (Power, 2005, Beaudry, 2016).

Much research on childhood deafness and hearing loss comes from the medical model (Geers et al, 2017; Grandon et al, 2019, Bruijnzeel et al, 2016) which views deafness as a sensory deficit or hearing impairment and investigates how subsequent medical interventions can improve hearing and remediate spoken language development delays. From many studies of DHH infants within a medical framework, several factors that contribute towards typical spoken language development in DHH children have been identified. These include, however are not limited to: early identification and aiding (Spencer & Marschark, 2010), age at implantation (Dettman et al., 2016), non-verbal cognitive ability (Cejas et al., 2018), underlying causes of deafness, and family involvement (Watkin et al., 2007).

A main topic in DHH infant language development research in the medical model is how to optimize spoken language development through earlier age of implantation (e.g. Bruijnzeel et al, 2016). Medical interventions attempt to restore functional hearing at a young enough age to reduce the risk of language delays. Decreases in children's age at implantation and increases in duration of CI usage are both associated with gains in language development (Bruijnzeel et al, 2016). However, it is currently not possible to implant DHH babies until around 9 months of age and in many countries, CI happens around 24 months. In the UK for example the selection process for CI candidacy and the ensuing time for medical appointments can delay the process to around 18 months. This leaves a significant time period, where DHH children have reduced access to spoken language. Thus, the medical model attributes part of the variability seen in DHH infants' language development after CI to the early period of reduced access to sound and spoken language.

In stark contrast to a medicalization of deafness, the social model looks at the whole DHH child, not just their deafness. A recurrent theme in this model is the acceptance and celebration of the acquisition of signed language by DHH children (Hall, Eigsti, Bortfeld Lillo Martin, 2018). A comparatively smaller field of research produced from the social model maintains that DHH children should not be 'medicalized' (Hall, Hall & Caseli, 2019) instead deafness should be viewed as an aspect of human diversity, indeed, rather than a hearing "loss," the social model refers to the importance of signed language, deaf culture and deaf "gain." The main evidence for language development and the importance of signed language within the social model comes from two sources: studies of deaf adults who learned signed language at deaf schools (before the implementation of neonatal screening, more children had severe language delays and were exposed to signing in the many deaf schools that existed) and language development of native signers. A native signer is a DHH child of DHH parents who is exposed to a signed language from birth. Native signers make up 5-10% of the DHH child population (Mitchell & Karchmer, 2006). The other 90-95% are infants with hearing parents who have no experience of signed language and deafness. Research in the social model argues that if native signers have 'normal' language and cognitive development, other DHH infants could achieve this if their parents learned to sign early and at a well enough level (e.g. Caseli, Pyers & Lieberman, 2021). Explanations for variability in language outcomes by researchers in the social model thus come from the environment rather than the child's deafness. Hearing and speaking parents who offer poor or no access to a signed language are 'depriving' the DHH infant of their 'natural' language (Hall et al., 2019; Davidson, Lillo-Martin & Pichler, 2014).

Interim Summary

DHH children's language development is influenced by factors originating from both within and surrounding the child. DHH babies are medically diagnosed deaf and receive early interventions focusing on their hearing via clinical professionals who prioritize hearing and speech. Some DHH children and their families meet with professionals, families, and other DHH adults and children, who provide experiences of signed language and what deafness means from within a social model. Researchers in deafness and language development from within each model put emphasis on different factors, that is, (1) hearing loss leads to reduced early access to environmental speech and variability in language development (medical model) and (2) poor access ('deprivation') and slow development of signed language leads to language delays (social model). It is possible that for a percentage of DHH children, both frameworks can be implemented, while neither alone is the answer for all DHH children. In moving the field forward, a set of issues in the literature need further investigation. The rest of this paper covers some of these issues.

Issues to clarify in the literature before moving forward

The CI ensures language development

The rapid increases in the number of DHH infants implanted and the improvement in CI technology has led to great advancements in spoken language development for most DHH children but not all (approximately 30% have continued language delay: Bruijnzeel, Ziylan, Stegeman, Topsakal & Grolman, 2016). Research studies from the medical model report that these problems are also apparent in more complex areas of language and during more demanding cognitive tasks (Geers, et al. 2009). Increased difficulties with higher-load tasks suggests that cognitive abilities supporting language development are also variable in DHH infants (Edwards & Isquith, 2020).

Hearing parents assume they can talk and interact with their DHH infant as they do with hearing babies (Marschark & Spencer, 2006). This assumption is natural, as most parents want their baby to be part of their own particular culture and social world. However, deafness alters the naturalness of this early interaction. While CI improve hearing, they cannot on their own solve the complex problem of how parents and children communicate best with each other. In hearing families, a particular early issue even after an early CI, is the establishment of successful communicative routines with the DHH infant (Levine, Strother-Garcia, Golinkoff & Hirsh-Pasek, 2016, Kelly et al, 2022).

Signed language delays spoken language development

Some studies written from the medical model suggest that learning signed language or sign support systems (e.g., Sign Supported English: SSE) can interfere with the development of spoken language (Geers et al., 2017), while other studies have found that learning signed language can in fact support the acquisition of spoken language (Davidson et al., 2014). An absence or severe reduction of spoken language input from parents (because they are signing all the time) would likely reduce the DHH child's ability to acquire spoken language. But this either/or situation is an unlikely one in the lives of DHH infants. It is more realistic that hearing parents speak and use signs together in some form of fluid bilingualism (Hermans, van Berkel-van Hoof & Knoors, 2021). More work is needed to determine the role of learning signed language or SSE as a facilitating factor for spoken language acquisition.

Deaf children experience language 'deprivation'

In the last few years, the pre-existing term 'language deprivation' has entered the deafness research literature in light of the social model (e.g. Caselli, et al 2021). 'Deprivation' is a very strong negative term in developmental psychology and especially language studies, stemming from work by Curtiss (2014) on Genie and the orphanage literature (Windsor, 2007). In the context of DHH children's development, is it accurate to talk about language deprivation at all? The first question we should ask is 'Who is doing the depriving'? The answer in the papers from the social model is hearing non-signing parents. Remember if we cannot look to the child's deafness as a cause of their language delay (that is the way the medical model sees deafness) we must look to the parents. However, many studies find no difference in the quantity of spoken language input to hearing and DHH infants (e.g. Ambrose, et al 2015).

It is theoretically more accurate to say hearing non-signing parents are 'depriving' the DHH child of signed language. However, aside from input the question is what does the infant uptake from the language addressed to them: what form of input is the most accessible? There have been very few empirical studies of hearing parents using signed language or SSE with their DHH infants. It is not clear what is optimal in terms of uptake: fluent and constant spoken language input, SSE or a learner version of signed language input (Lu, Jones & Morgan, 2016).

Generalizability of research on deaf children who are native signers

Much language development research in the social model uses a small group of DHH native signing children and reports age-appropriate early language and cognitive development (e.g. Wolfe, Herman, Roy & Woll, 2010). The explanation offered from studies of native signers for this optimal outcome is high quality input from DHH adults who use visual-tactile strategies during interaction, which offers potentially maximal uptake for DHH infants. Researchers often generalize these findings to the 90-95% of DHH infants with hearing parents. All DHH children, the argument goes, have as much potential to learn language given a specific language environment i.e. parents who are fluent users of a sign language. There remains an assumption in the social model that hearing parents can provide this type of language input and secondly DHH children can benefit from this input to develop age-appropriate language and cognitive abilities. There are insufficient studies of DHH children with hearing parents who use signs to answer these questions at this point (but see Caseli et al, 2021). More research on hearing parents using signs, as well as, SSE will allow language development investigators to support policy makers. Should we organize health-funded interventions so that all DHH children are exposed to signs (social model) or just focus sign interventions on those children who have less fluency with spoken language, despite all the medical interventions being made available (medical model)? This is a complicated question as hearing parents and professionals do not know, in the majority of cases, in advance if their DHH baby is going to be a successful spoken language learner (some DHH children have known factors mediating spoken language success e.g. malformed cochlea). The social model responds to this uncertainty with a straightforward answer: 'all DHH infants need to learn signed language just in case' (Wiefferink, et al, 2008). Yet, language learning as adults is not a trivial thing for parents in terms of effort, outcome and resources. At the same time these same parents and other children in the family are already native speakers of one or more spoken language. For the same question related to educational provision see Knoors

& Marschark (2012).

Common ground: communication as the goal rather than words or signs

Returning to one aspect highlighted previously as contributing to variability in language development: family involvement or quality of parent-child interaction. Morgan, Curtin & Botting (2021) highlighted one aspect of early communication development as being particularly vulnerable to developmental disruptions during the first 12 months of life: the establishment of intersubjectivity. Intersubjectivity refers to the establishment of meaningful and reciprocal exchanges between individuals. Intersubjectivity develops between the infant and the parent through contingent interaction during communication (Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008). There are several advantages for future language learning that stem from good early social interaction: DHH infants' ability to reciprocate, share attention and intentionally communicate with hearing family members. DHH infants who are engaged and attempt to communicate socially can also shape the hearing adult's responses. In this perspective, first the DHH infant grasps that the function of social interaction is to share ideas. Once this is in place, then the infant is motivated to learn the symbols (words, signs or SSE) to express and receive these ideas.

Rather than focusing on speaking or signing there is a compromise position that both models can work together to achieve. Both the medical and social models can agree that successful communication patterns, in whatever form or combination of language/s that the parent and child are comfortable using, if established early in development, will foster good future language development. There are a handful of intervention studies that take this compromise position with positive results (Roberts, 2018; Kelly et al, 2022).

Conclusion

Research over the last 30 years has documented great advancements in the language outcomes of DHH children following neo-natal screening and the implementation of early CIs. At the same time, there is a continuing debate as to why a proportion of DHH infants continue to display variability in outcomes. Explanations have been put forward stemming from a medicalized model of the DHH infant, stressing access to sound, and a social model which points to a lack of access to signed language as explaining variability. Both sides of this debate have taken standpoints concerning the CI as a cure for deafness, that signed languages impinges on spoken language, the 'language deprivation' DHH children experience from hearing parents and the generalizability of research on native signers. In all these areas it is vital that more research is carried out before any concrete recommendations can be offered. One area which offers a middle ground is around improving the quality of early parent-child interaction and the establishment of intersubjectivity to support future cognitive and language development.

References

Beaudry J. S. (2016). Beyond (Models of) Disability?. *The Journal of Medicine and Philosophy, 41*(2), 210–228. https://doi.org/10.1093/jmp/jhv063

Boonen, N., Kloots, H., Nurzia, P., & Gillis, S. (2021). Spontaneous speech intelligibility: Early cochlear implanted children versus their normally hearing peers at seven years of age. *Journal of Child Language*, 1-26. doi:10.1017/S0305000921000714

Bornstein, M., Tamis-LeMonda, C., Hahn, C.-S., & Haynes, M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular and specific parental construct. *Developmental Psychology, 44*(3), 867–874. https://doi.org/10.1037/0012-1649.44.3.867.

Bruijnzeel, H., Ziylan, F., Stegeman, I., Topsakal, V., & Grolman, W. (2016). A Systematic Review to Define the Speech and Language Benefit of Early (<12 Months) Pediatric Cochlear Implantation. *Audiology and Neurotology, 41*(2), pp. 113–126. https://doi.org/10.1159/000443363

Cassani G, Grimm R, Daelemans W,. & Gillis, S. (2018). Lexical category acquisition is facilitated by uncertainty in distributional co-occurrences. *PLoS ONE 13*(12): e0209449. https://doi.org/10.1371/journal.pone.0209449

Caselli, N., Pyers, J. & Lieberman, A.M. (2021). Deaf Children of Hea*ring Parents Have Age-Level Vocabulary Growth When Exposed to American Sign Language by 6 Months of Age,* The Journal of Pediatrics, 232, 229-236

Cejas, I., Mitchell, C. M., Hoffman, M., Quittner, A. L., Santina, C. della, Marsiglia, D., Martinez, D., Telischi, F., Glover, R., Sarangoulis, C., Zwolan, T., Arnedt, C., Teagle, H. F. B., Woodard, J., Eskridge, H., Eisenberg, L. S., Johnson, K., Fisher, L., Ganguly, D. H., Wang, N. Y. (2018). Comparisons of IQ in children with and without cochlear implants: Longitudinal findings and associations with language. *Ear and Hearing, 39*(6), 1187–1198. https://doi.org/10.1097/AUD.0000000000000578

Coene, M., Schauwers, K., Gillis, S., Rooryck, J., & Govaerts, P. (2011) Genetic predisposition and sensory experience in language development: Evidence from cochlear-implanted children. *Language and Cognitive Processes, 26*(8), 1083-1101

Curtiss S. (2014). *Genie: A psycholinguistic study of a modern-day wild child.* New York: Academic Press.

Davidson, K., Lillo-Martin, D., & Pichler, D. C. (2014). Spoken English language development among native signing children with cochlear implants. *Journal of Deaf Studies and Deaf Education, 19*(2), 238–250. https://doi.org/10.1093/deafe d/ent045

Dettman, S. J., Dowell, R. C., Choo, D., Arnott, W., Abrahams, Y., Davis, A., Dornan, D., Leigh, J., Constantinescu, G., Cowan, R., & Briggs, R. J. (2016). Long-Term communication outcomes for children receiving cochlear implants younger than 12 months: A multicenter study. *Otology and Neurotology, 37*(2), e82–e95.

https://doi.org/10.1097/MAO.0000000000000915

Edwards, L. C., & Isquith, P. K. (2020). Cognitive Development. In Marc Marschark & H. Knoors (Eds.), *The Oxford Handbook of Deaf Studies in Learning and Cognition* (pp. 197–212). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190054045.013.7

Odijk, L. & Gillis, S. (2022). Parents tune their vowels to the emergence of children's words. Journal of Child Language, 1-20. Grandon, B., Vilain, A. & Gillis, S. (2019). The effect of hearing impairment on the production of prominences: The case of French-speaking school-aged children using cochlear implants and children with normal hearing. *First Language, 39*, 200-219.

Geers, A. E., Moog, J. S., Biedenstein, J., Brenner, C., & Hayes, H. (2009). Spoken language scores of children using cochlear implants compared to hearing age-mates at school entry. *Journal of Deaf Studies and Deaf Education, 14*(3), 371–385. https://doi.org/10.1093/deafed/enn046

Geers, A. E., Mitchell, C. M., Warner-Czyz, A., Wang, N.-Y., Eisenberg, L. S., & CDaCl Investigative Team. (2017). Early sign language exposure and cochlear implantation benefits. *Pediatrics, 140*(1).

Hall, M. L., Eigsti, I.-M., Bortfeld, H., & Lillo-Martin, D. (2018). Executive function in deaf children: Auditory access and language access. *Journal of Speech, Language, and Hearing Research, 61*, 1970-1988.

Hall, M.L., Hall, W.C., & Caselli, N.K. (2019). Deaf children need language, not (just) speech. *First Lang 39*:367–395.

Hoffman, M.F., Cejas, I. & Quittner, A.L. (2016). The Childhood Development after Cochlear Implantation (CdaCI) Investigative Team. Comparisons of longitudinal trajectories of social competence: Parent ratings of children with cochlear implants versus hearing peers. *Otol Neurotol 37*:152–159.

Kelly, C.; Crawford, E.; Morgan, G.; McKinley, D. Matthews, D. (2022). Assessing a Video-Based Intervention to Promote Parent Communication Strategies with a Deaf Infant: A Feasibility and Acceptability Study. *J. Clin. Med. 2022, 11*, x. https://doi.org/10.3390/xxxxx

Knoors, H. & Marschark, M. (2012). Language Planning for the 21st Century: Revisiting Bilingual Language Policy for Deaf Children. *The Journal of Deaf Studies and Deaf Education, 17*, 291–305, https://doi.org/10.1093/deafed/ens018

Levine, D., Strother-Garcia, K., Golinkoff, R.M., & Hirsh-Pasek, K. (2016). Language development in the first year of life: what deaf children might be missing before cochlear implantation. *Otol Neurotol 2016*; 37:56–62.

Lu, J., Jones, A., & Morgan, G. (2016). The impact of input quality on early sign development in native and non-native language learners. *Journal of Child Language, 43*(3), 537–552. https://doi.org/10.1017/S0305000915000835.

Marks, D. (1997). Models of disability. *Disability and Rehabilitation, 19*:3, 85-91, DOI: 10.3109/09638289709166831

Marschark, M., & Spencer, P. E. (2006). Spoken Language Development of Deaf and Hard-of-Hearing Children: Historical and Theoretical Perspectives. In P. E. Spencer & M.

Marschark (Eds.), *Advances in the spoken language development of deaf and hard-of-hearing children* (pp. 3–21). Oxford University Press.

Mitchell, R.E. & Karchmer, M. (2004). Chasing the mythical ten percent: parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies 4*:138-63.

Morgan, G., Curtin, M., & Botting, N. (2021). The interplay between early social interaction, language and executive function development in deaf and hearing infants. *Infant behavior & development, 64*, 101591. https://doi.org/10.1016/j.infbeh.2021.101591

National Institute for Health and Clinical Excellence (NICE). (2019). Cochlear implants for children and adults with severe to profound deafness. *NICE technology appraisal guidance 566*. [online]

Power, D. (2005). Models of Deafness: Cochlear Implants in the Australian Daily Press, *The Journal of Deaf Studies and Deaf Education, 10*(4), 451–459. https://doi.org/10.1093/deafed/eni042

Roberts, M. (2018). Parent-Implemented Communication Treatment for Infants and Toddlers with Hearing Loss: A Randomized Pilot Trial. *Journal of Speech Language and Hearing Research. 62*. 1. 10.1044/2018_JSLHR-L-18-0079.

Spencer, P., & Marschark, M. (2010). *Evidence of best practice models and outcomes in the education of deaf and hard-of-hearing children: an international review.* Oxford University Press.

Theunissen S, Rieffe C, Kouwenberg M, et al. Depression in hearing-impaired children. *Int J Pediatr Otorhinolaryngol 75*:1313–1317.

Vaccari C. & Marschark, M. (1997). Communication between parents and deaf children: implications for social-emotional development. *J Child Psychol Psychiatry 38*:793–8012.

Watkin, P., McCann, D., Law, C., Mullee, M., Petrou, S., Stevenson, J., Worsfold, S., Ho, M. Y., & Kennedy, C. (2007). Language ability in children with permanent hearing impairment: The influence of early management and family participation. *Pediatrics, 120*(3). https://doi.org/10.1542/peds.2006-2116

Wiefferink, C., Spaai, G., Uilenburg, N., Vermeij, B., & De Raeve, L. (2008) Influence of Linguistic Environment on Children's Language Development: Flemish Versus Dutch Children. *Deafness & Education International, 10*:4, 226-243. DOI: 10.1179/146431508790559724

Windsor, J., Glaze, L. & Koga, S. (2007). Language Acquisition with Limited Input: Romanian Institution and Foster Care. *Journal of Speech, Language, and Hearing Research. 50*. 1365-81. 10.1044/1092-4388(2007/095).

Woolfe, T., Herman, R., Roy, P., & Woll, B. (2010). Early vocabulary development in deaf native signers: a British Sign Language adaptation of the communicative development inventories. *Journal of Child Psychology and Psychiatry, 51*(3). https://doi.org/10.1111/j.1469-7610.2009.02151.x

STEM-, SPRAAK- EN TAALPATHOLOGIE

Gepubliceerd online: 5 november 2022

University of Groningen Press
Vol. 27, 2022, Special Issue, pp. 145-176
https://doi.org/10.21827/32.8310/2022-SG-145
Special Issue for Steven Gillis

Auditory brainstem implantation in children: the case of place of articulation

Lotte Odijk & Jolien Faes

Computational Linguistics & Psycholinguistics (CLiPS) Research Center, University of Antwerp

Samenvatting

Deze studie onderzoekt de uitspraak van de articulatieplaats van spraakklanken in de spontane spraak van drie kinderen met een herstenstamimplantaat. Deze kinderen worden vergeleken met twee controlegroepen: kinderen met een cochleair implantaat en kinderen met een normaal gehoor. Het spraakmateriaal werd betrokken uit de spontane spraak tussen de kinderen en hun zorgverleners. Voor elke doelklank werd de plaats van articulatie gecodeerd en de plaats van articulatie zoals die door de kinderen werd gerealiseerd. Plaats van articulatie werd gecodeerd als labiaal, coronaal of dorsaal. De resultaten tonen aan dat "coronaal"het vaakst voorkwam in de eigen producties. Eén kind vormde hierop een uitzondering en had een voorkeur voor "labiaal". De labiale plaats van articulatie werd ook accurater uitgesproken dan de coronale plaats van articulatie. De dorsale plaats van articulatie was het minst accuraat. Kinderen met een hersenstamimplantaat waren het minst accuraat. Als de plaats van articulatie niet juist werd uitgesproken, dan werd de klank vaak weggelaten en niet vervangen door een klank met een andere plaats van articulatie. De algemene conclusie is dat kinderen met een hersenstamimplantaat wel degelijk baat hebben bij een implantaat, maar dat zij nog een hele weg af te leggen hebben om achterstand in te lopen. Het gebruik van gebarentaal is nuttig om de communicatie te ondersteunen.

Abstract

This study aimed to investigate the development of the production of place of articulation of three children with auditory brainstem implants (ABI) in spontaneous speech production. The main participants were three children implanted with an ABI. They were compared against two different control groups: children with a cochlear implant (CI) and children with normal hearing (NH). Data was obtained from spontaneous speech between the children and their caregivers. For each word production, the place of articulation of both the target word as the child's own production was identified. This was broadly identified in three categories: labial, coronal and dorsal. The analysis revealed

Correspondentieadres:
Lotte Odijk
University of Antwerp
Computational Linguistics and Psycholinguistics (CLiPS) Research Center
Prinsstraat 13, 2000 Antwerpen, Belgium
E-mail: lotte.odijk@uantwerpen.be

that in general, the coronal place of articulation was most used in all children's own production, as well as in the target words, except for one child with ABI, who showed a preference for labials. In terms of accuracy, labial place of articulation was produced more accurately than the coronal place of articulation for all children. The dorsal place of articulation had very low accuracy probabilities. Children with ABI had the lowest accuracy rates. When the place of articulation was not correctly produced, they were often omitted instead of replaced by another place of articulation. It was concluded that the children with ABI benefit from their device, but still have a long way to go to catch up to their peers. It is suggested that sign language is needed to guarantee smooth communication.
Keywords: auditory brainstem implantation; pediatric; oral language, place of articulation; labial; coronal; dorsal

Introduction

This study reports on the development of place of articulation in three children with auditory brainstem implants (ABI), in comparison to a group of children with cochlear implants (CI) and a group of children with typical, normal hearing (NH).

In the past several decades, different techniques have been emerging to restore hearing capacity in patients with a severe-to-profound sensorineural hearing loss. The decision for one or the other therapeutic aid is usually determined by the type and/or locus of the hearing deficit. With respect to sensorineural hearing loss, there are two implantable aids: a CI and an ABI. With experience of adult patients, both techniques have been expanded to pediatric populations as well. CIs have been implanted in children in the last two decades of the previous century. In contrast, ABI has only been emerging in pediatric hearing restoration since the beginning of this century, with the first implantation in 2001 (Colletti et al., 2001).

A CI is used when the hearing loss results from absent or malformed hair cells in the cochlea, whereas an ABI is used when the hearing loss results from an absent auditory nerve or malformed or ossified cochlea, precluding a CI placement. Studies recommend ABI implantation only if CI placement is anatomically impossible or after a period of CI use with poor language outcomes (Batuk et al., 2020; Buchman et al., 2011; Farhood et al., 2017; Hammes Ganguly et al., 2019). The external part of both devices, ABI and CI, consists of a microphone that captures the environmental sounds and a processor that transforms them into a digital code. The internal part consists of an electrode array, but the placement of this array depends on the type of implant. In a CI, this electrode array is inserted into the cochlea, preserving the tonotopic organization of the cochlea and directly stimulating the auditory nerve. In an ABI, this electrode array is inserted directly onto the cochlear nucleus of the brainstem, bypassing the entire cochlea and the auditory nerve. In contrast to the CI – which maintains the tonotopic organization-, the organization of hearing pathways of the brainstem are unclear and unpredictable in ABI, which seems to have lower hearing benefits as a result (Long et al., 2005; Wong et al., 2019). Yet, some children wear both a CI and a contralateral ABI. The first research results seem to suggest that this CI-ABI combination enhances the effect of both implants (Batuk et al., 2020; Friedman et al., 2018), although clearly more research is needed.

Children with ABI

Children with a congenital severe-to-profound hearing loss reach hearing thresholds between 30 and 60 dB HL (decibels hearing level) after ABI implantation (Colletti et al., 2004; Sennaroglu, Colletti, et al., 2016; Teagle et al., 2018; Yucel et al., 2015). The ABI offers sound awareness, sound discrimination and identification of phonetic contrasts to these children. Research has pointed out that speech perception outcomes are better when children are implanted earlier (Aslan et al., 2020), when children have lower hearing thresholds after their ABI implantation (Sennaroglu, Sennaroglu, et al., 2016; Yucel et al., 2015), and when children have no additional disabilities (Colletti et al., 2014; van der Straaten et al., 2019). The children who meet these three specifications can at least understand easy phrases without the aid of lip reading. Nevertheless, there are also considerable individual differences in their open set speech perception outcomes, even when they met these specifications (Colletti et al., 2014; Colletti et al., 2004; Sennaroglu, Colletti, et al., 2016).

Children who achieve open speech perception (i.e., the children with low hearing thresholds after early implantation and without other disabilities) also develop speech production skills. This development follows the pathway of typical spoken language development, in the sense that these children develop vocalizations, babble, and start to produce words and sentences after a period of ABI experience (e.g. Bayazit et al., 2014; Faes & Gillis, 2018, 2019a, 2019b) as children with typical development. Nonetheless, the pace of development is very slow (Aslan et al., 2020; Eisenberg et al., 2018; Faes & Gillis, 2019b, 2021a; Teagle et al., 2018; van der Straaten et al., 2019) and there is considerable variation between the children with ABI. With increased hearing experience, children with ABI use basic word patterns (Eisenberg et al., 2018; Faes & Gillis, 2021b) in their increasing lexicon (Faes & Gillis, 2019b). In their speech production, language ambient phonemes and words appear after two to three years of device use, even though the accuracy of these phonemes and words is still low (Eisenberg et al., 2018; Faes, Gillis, et al., 2022; Faes & Gillis, 2021a; Teagle et al., 2018).

In sum, children with ABI with open set speech perception develop speech production, but this development is slow. When compared to children with cochlear implants (CI) – another group of children with congenital severe-to-profound hearing loss – and children with typical, normal hearing (NH), children with ABI clearly lag behind even after five to six years of device use. For instance, the best performing children with ABI have expressive language skills that are to be situated between those of children with CI with and without additional disabilities (van der Straaten et al., 2019). These best performing children with ABI were those early implanted children with low hearing thresholds and no additional disabilities. Going into detail, several aspects of lexical and phonological development of such best performing children with ABI lag behind those of children with CI and children with NH. For instance for lexical development, children with ABIs vocabulary sizes fall out of the lower border of the 95% confidence intervals of the children with CI and NH (Faes & Gillis, 2019b). For phonological development, the same is found for speech production accuracy and phonological complexity in speech: these fall out of the lower borders of the 95% confidence interval in children with CI and NH (Faes & Gillis, 2021b). In line with these results, the overall speech intelligibility of children with ABI was scored considerably lower than that of children with CI and NH, even after more than three years of device use (Faes, De Maeyer, et al., 2022). Overall, it takes children with ABI five to six years of device use in order to be intelligible for a familiar listener with or without lipreading (Aslan et al., 2020; Sennaroglu, Sennaroglu, et al., 2016; van der Straaten et al., 2019).

Effect of hearing loss on place of articulation

The important question is whether children with ABI reach similar spoken language proficiency as their CI or even NH peers. The current study will extend on the question by focusing on the phonemic accuracy of the children and specifically place of articulation. Phonemic accuracy is essential for children's speech intelligibility (Ingram, 2002). In general for children with NH, labial and coronal place of articulation are acquired before dorsal place of articulation (Beers, 1995). However, not much is known about the occurrence and the accuracy of place of articulation of children with ABI. One study has shown the consonant inventories of children with ABI (Faes & Gillis, 2021a), but here, the set off levels to include a certain consonant into a phonemic inventory were set at 2 occurrences and/or 50% or 75% accuracy rate. Even though this gives a good idea of the developmental trend, this does not give insight into the precise accuracy of production. Still, it seems that there was not much of a pattern to be found over the different children, rather all children showed highly individual patterns of in their developing phonemic inventory.

More research has been done on the performance of children with CI. Research to non-word repetition found that children with CI produced coronals with greater accuracy than labials and dorsals. They tended to replace labial and dorsal targets with coronals (Dillon, Cleary, et al., 2004; Dillon, Pisoni, et al., 2004; Moreno-Torres & Moruno-López, 2014). This result is consistent with the theory of Cummings et al. (2020), which states that the coronal place of articulation is the default place of articulation, and that the labial and dorsal place of articulation only appear with maturation and exposure to language. It is also possible that the poorer performance on labials is due to the absence of visual cues in the study design of these non-word repetition tasks, such as lip closure (Dillon, Pisoni, et al., 2004). Evidence for this can be found in research on children's use of consonants in spontaneous speech. Warner-Czyz and Davis (2008) showed that children with CI and NH differed on the frequency of consonants and consonant accuracy before the age of 24 months. Children with NH produced labial and coronal consonants with equal frequency, while children with CI produce more labials than coronals. Both groups produced dorsals the most infrequently. The children with NH produced consonants more accurately than children with CI. Children with CI used more omission, but their performance gradually increased over time. The finding that children with CI use more labial consonants than children with NH is not surprising. It is found that less proficient children with CI rely more on visual cues than auditory cues, and lipreading is even essential for proficient CI users (Huyse et al., 2013). Since the labial place of articulation is visually the most prominent, this would be easier to learn.

Because children with ABI are often implanted later in life than children with CI, sign language is an important part of their communication (Faes & Gillis, 2019a). It has been shown that hearing impaired children with a cochlear implant who are educated using both signed and spoken language, have poorer speech perception skills than hearing impaired children who have been educated using only spoken language (Bergeson et al., 2003). Furthermore, later implanted children with CI rely more on visual cues for speech perception than auditory cues (Bergeson et al., 2005). Likewise, for instance Schorr et al. (2005) and Most et al. (2009) showed that children with CI highly rely on visual cues in their speech perception, especially in challenging auditory conditions. Moreover, children with CI rely more on visual cues than NH-listeners do (Desai et al., 2008). So, it could be inferred that children with ABI, who also have poorer speech perception skills, also need more visual cues. As speech production depends on speech perception (Altvater-Mackensen & Fikkert, 2010; Jusczyk, 1992; Stoel-Gammon, 2011; Stoel-Gammon & Sosa, 2007), this could have consequences for their performance relating to place of articulation.

The present study

In the present study, the comparison between children with ABI and the control groups, viz. children with CI and children with NH, will be held against a language-intrinsic comparison measure, namely level of lexical development. More specifically, children with a similar amount of cumulative vocabulary in their spontaneous speech will be compared to one another. By using lexical development as a yardstick in the comparison, some age-related issues can be avoided. Children with ABI, as well as children with CI, have a delayed onset of hearing, so it seems inappropriate to compare them to children with NH on chronological age. Yet, a frequently used alternative option is to compare children in terms of hearing experience (see for instance Blamey, Barry, Bow, et al., 2001; Blamey, Barry, & Jacq, 2001; Eriks-Brophy et al., 2013; Faes & Gillis, 2019a; Schauwers, 2006; Szagun & Stumper, 2012). However, even in that case, there are issues with later implantation of the children with ABI as compared to the children with CI. In that case, the chronological age difference still has an influence, even in children with a similar amount of hearing experience, so that the use of hearing age as a comparative measure in this study also seems suboptimal. Therefore, lexical development was chosen as a language-intrinsic measure in the present study. This measure is increasingly being used to compare children with CI and ABI. Lexical development has been put forward in previous research as it has been shown to closely follow phonological development in children with CI and NH (e.g. Faes & Gillis, 2016; Smith et al., 2006; Sosa & Stoel-Gammon, 2012; van den Berg, 2012), and a similar trend seems to emerge in research involving children with ABI (Faes & Gillis, 2021a). In addition, lexical development is more related to phonological development than chronological age (van den Berg, 2012), so that it seems the least skewed measure of comparison.

Place of articulation is operationalized in three categories: labial, coronal and dorsal in line with Booij (1995)'s division of the place feature (Figure 2.3 on page 9 in Booij (1995). Also in Booij (1995), the Dutch consonant system is depicted: Labial consonants are bilabial and labio-dental consonants, coronals are alveolar and palatal consonants and dorsals are velar consonants. The distinction between these categories is one of the most common contrasts in different languages. In terms of token frequencies, coronal consonants appear the most frequently in Dutch (Luyckx et al., 2007).

Four research questions will be assessed in the present study:
(RQ1) Occurrence of the different places of articulation in target: how often are the three different places of articulation targeted by the children? What is the distribution in the target words?
(RQ2) Occurrence of the different places of articulation in replica: how often do children use the different places of articulation in their own productions – regardless their correctness?
(RQ3) Accuracy of place of articulation: how accurately is place of articulation produced by the children?
(RQ4) Error analyses: if the place of articulation was produced incorrectly, is it replaced or omitted? And, if it is replaced, which other place of articulation?

For each of these research questions, a comparison between each child with ABI and the control groups (children with CI and children with NH) will be carried out. As mentioned earlier, the comparison will be based on similar lexicon sizes. There are two possibilities with respect to the four research questions: (a) children with ABI produce place of articulation differently as compared to children with CI and children with NH, (b) children with ABI show a similar pattern in place of articulation as compared to children with CI and children with NH. With respect to the first possibility, there seem to be two valid hypotheses according to the literature.

A first possibility is that there is a difference in labial place of articulation in ABI children as compared to the control groups. For children with ABI, as well as children with CI, it has been shown that the speech signal provided by their device is degraded and more noisy than in normal hearing (Drennan & Rubinstein, 2008). The literature seems to indicate that the speech signal is even more degraded in ABI than in CI (Wong et al., 2019). In addition, children with ABI are generally implanted later than children with CI. So, sign language is more common in ABI children than in children with CI (and by extension also children with NH). This is due to their later implantation, the need for a different communication mode before implantation and the more degraded speech ABI signal, resulting in a greater need for supportive sign language for efficient communication. Since the labial place of articulation is visually the most salient and children with ABI are more used to pick up visual cues for speech perception, a more prominent role of labial consonants in children with ABI's speech productions is expected. In other words, it is expected that labials are more present in children with ABI's word productions (replica, RQ2), that they are more accurate in children with ABI (RQ3) and that they are used as an alternative in erroneous productions (RQ4). For the target words, RQ1, there might also be an effect of children with ABI targeting words with a labial place of articulation, because of their own increased attention for it. This result would be in line with the results of Warner-Czyz and Davis (2008) with respect to comparison CI-NH.

A second possibility is that children with ABI produce more and more accurate coronal place of articulation as compared to the control groups. In adult speech production, it is evident that the coronal place of articulation is the default place of articulation, also called underspecified place of articulation (see e.g. for an review in the introduction of Cummings et al., 2020). Cummings et al. (2020) proposed that acquiring place of articulation is a developmental process, in which first the default [coronal] is acquired, as it is the least complex. Later on, more complex – specified – place of articulation, i.e. labial and dorsal, appear as well. Cummings et al. (2020) indicated that time and, especially, exposure to language are needed to develop place of articulation. For instance, they suggested that four-to-six-year-old children with phonological disorders have less fine-grained phonological representations, as compared to a typically developing group of peers, resulting in a less mature stage of phonological acquisition, so that they had only acquired coronal place of articulation. Extrapolating this to children with ABI, a similar pattern may present itself. Due to a degraded speech signal, the phonological representations of children with ABI are highly likely the be less fine-grained than in children with CI and children with NH. Faes and Gillis (In press), for instance, showed significantly greater variability in word production in children with ABI than in children with CI, which can be explained by less-fine-grained phonological representations in the ABI group. In addition, children with ABI have had less exposure to language as compared to the control groups. As a result, it is possible that children with ABI are in an earlier stage of development like children with phonological disorders in Cummings et al. (2020). This would result in a more extended use of the coronal place of articulation in children's productions and error patterns as well as a more accurate production of this place of articulation. This result would be in line with Dillon, Cleary, et al. (2004); Dillon, Pisoni, et al. (2004); and Moreno-Torres and Moruno-Lopez (2014) for the CI – NH comparison.

Methods

Participants

The present study reports on the speech production of three children with ABI. In the period 2015 to 2019, only eight children received an ABI in Belgium. Inclusion criteria for the present study were: a congenital bilateral severe-to-profound hearing loss, no other health or developmental problems, growing up in a Dutch-speaking family and having completed at least one and a half years of follow-up. These three of children had been followed up longitudinally on a monthly basis. The study was approved by the Ethical Committee for the Social Sciences and Humanities of the University of Antwerp (EASHW_16_29) and all parents signed an informed consent.

ABI1 and ABI2 were born with a sensorineural hearing loss, due to the absence of their auditory nerves. Their pure tone average (PTA) hearing thresholds before implantation were 120 dB HL and 116 dB HL respectively. Both children were implanted around their second birthday (24 months and 25 months of age), with a Med-El ABI (Synchrony and Concerto respectively). Two months after surgery, the ABI was fitted. In both children nine of the 12 electrodes were activated. Two years after ABI surgery, pure tone average thresholds had improved to 37.5 dB HL and 43 dB HL. ABI1 was bilaterally implanted, this child received a second, contralateral ABI at four years and nine months of age. Longitudinal monthly data collection started about one year after implantation for ABI1 and two years after implantation for ABI2 and lasted more than two years in both children. To be precise, data collection for ABI1 started when the child had three years and two months of age and continued up till the age of five years and seven months. The data collection of ABI2 started at four years and one months of age and continued up till six years and three months of age. Both children were raised with oral communication, supported by Flemish Sign Language. The amount of sign language in the input was larger for ABI1 than for ABI2.

ABI3 first received a CI at eight months of age, due to a severe-to-profound hearing loss after a diagnosis of auditory neuropathy. The child's PTA hearing threshold was 92.5 dB HL in the better, CI-implanted ear. With CI, the mean PTA improved to 33 dB HL. Nevertheless, language and communication skills did not develop well. Therefore, a contralateral ABI was implanted at the age of four. Two months after implantation, all electrodes could be fitted. The data collection for this child started two months before ABI implantation (at three years and ten months of age) and went on for one year and a half, up till five years and four months of age. ABI3 was also raised in oral Dutch, supported by Flemish Sign Language (similar amount of sign language input as ABI1).

Two control groups were included in the study: (a) children with a cochlear implant, and (b) children with typical hearing. Nine children with CI participated in the study as a first control group (Table 1). These children were born with a sensorineural hearing loss with no other health or developmental problems. Their sensorineural hearing loss was associated to underlying pathologies different from those in ABI children, causing deficits in their cochlea. The mean PTA before implantation was 112.56 dB HL. All children were implanted well before their second birthday, with a range of 5 to 20 months and a mean implantation age of one year (SD = 5 months). After implantation, the mean PTA hearing threshold was 32.22 dB HL at the age of two. Six children received a bilateral implant at a later age (Table 1). All children were raised in spoken Dutch, with only a limited number of lexical signs in support. Longitudinal monthly data collection started after implant fitting and went on for 30 months after implantation (and yearly thereafter).

Thirty children with NH were selected as a second control group. These children had no reported health, hearing, or developmental problems. All children were raised in spoken Dutch and were followed monthly between six months and two years of age.

Data collection and transcription

Monthly spontaneous video recordings were made at the children's home, capturing spontaneous conversations between the children and their caregiver(s). These video recordings were entirely unstructured (participants were free to do whatever they wanted), so that they resembled daily life interactions. All recordings lasted about one hour. For the control groups, these one-hour video recordings were reduced to 20-minute selections, to keep transcription time within reasonable limits. For detailed information, see Molemans (2011),

Table 1: Individual characteristics of children with CI

ID	Gender	PTA unaided (dB HL)	PTA CI (dB HL) (age 2;00)	Age CI implantation	Age second CI
CI1	F	120	48	1;01	6;03
CI2	F	120	30	0;07	4;08
CI3	F	115	33	0;10	5;10
CI4	M	113	48	1;06	-
CI5	M	93	38	1;05	6;04
CI6	M	120	53	0;09	-
CI7	F	117	42	0;05	1;03
CI8	F	112	38	1;07	-
CI9	F	103	28	0;08	1;11

dB HL = decibel Hearing Level. Ages are presented in years;months. - = no second CI.

Schauwers (2006), van den Berg (2012) and Van Severen (2012). Because of the smaller number of children in the ABI group, ABI video recordings were transcribed entirely.

The video recordings were transcribed in CHILDES' CLAN following the CHAT conventions (MacWhinney, 2000). All children's lexical productions were transcribed orthographically and phonemically. In addition, a phonemic transcription of the target word was added to these transcriptions, using the Flemish pronunciation database Fonilex (Mertens, 2001). For the ABI group, interrater reliability equalled 79.90% (SD = 3.57) in a phoneme-to-phoneme comparison. For the CI and NH group, interrater reliability for consonant place of articulation was 82.90% and 81.14% respectively (for more information on data reliability assesment, see methods sections of Faes, 2017; Molemans, 2011; Schauwers, 2006; van den Berg, 2012; Van Severen, 2012).

Data analyses and statistical approach

The cumulative vocabulary was counted for each child in each group. That is, in the transcription of the first speech sample of a child, all distinct word types were tallied, constituting the vocabulary count at this point. Then, in the transcription of consecutive speech samples, the vocabulary size was increased each time a new distinct word type appeared in the transcription. For ABI1, the cumulative vocabulary size went up to 450 word types, for ABI2 it varied between 50 and 650 word types and for ABI3 up to 350 word types. In order to match the data, only CI and NH files with corresponding cumulative vocabulary counts were selected in the analyses.

For each word production, the place of articulation of both the target word and the child's own production (henceforth replica) was identified for singleton consonants in the target. Place of articulation was broadly classified in three categories: labial, coronal or dorsal. With respect to children's replicas, a fourth option 'omitted' was added if the child did not produce the target phoneme. For instance, for the target word /buk/ *boek* (Eng. book) with a replica /tu./ (with the dot representing an omitted phoneme) the places of articulation for the target /b/ and /k/ were determined (labial and dorsal) as well as the places of articulation for the child's production, i.e. the replica /t/ and /./ (coronal and omitted).

The four research questions were investigated in the following way:
(1) Occurrence of the different places of articulation in targets: how often are the three different places of articulation targeted by the children? What is the distribution in the target words?
(2) Occurrence of the different places of articulation in replicas: how often do children use the different places of articulation in their own productions - regardless their correctness?
(3) Accuracy of places of articulation: how accurately are the places of articulation produced by the children?
(4) Error analyses: if production of place of articulation incorrect, is it replaced or omitted? And, if it is replaced, by which other place of articulation?

For research questions (1), (2) and (4), multinomial logistic regression analysis was carried out in R using the multinom function from *nnet* package (Venables & Ripley, 2002). The fixed effects in each of these models were Hearing status (CI and NH) and the effect of cumulative vocabulary. Interactions between these effects were considered as well. For research question (3), a binomial generalized logistic regression was used (glm). The fixed effects were Hearing status (CI and NH), cumulative vocabulary, place of articulation in the target (PoA_Target) and the interaction between Hearing status and place of articulation in the target (PoA_Target). The intercept was set at a cumulative vocabulary size of 100 word tokens for all analyses. For sake of clarity, the log odds from the models were retransformed to probabilities in the figures.

Results

Frequency distribution in children's speech

In order to study the occurrence of the different places of articulation in children's productions, two analyses were carried out. A frequency distribution was drawn with respect to the three places of articulation in children's target words (labial, coronal, and dorsal) and a frequency distribution was drawn for the places of articulation in their own productions, i.e. replicas (labial, coronal, dorsal, and omitted). The results of these analyses can be found in Figures 1 and 2 respectively and Tables A1 and A2 in the Appendix.

In the target words of all children (ABI1, ABI2, CI and NH), except ABI3, the coronal place of articulation is the most likely, followed by the labial place of articulation and the dorsal one. For ABI3, however, the labial place of articulation occurs most frequently in the target, even though there is a decrease with increasing lexicon size. These patterns can be clearly observed in Figure 1, and the differences between the places of articulation for each (group of) child(ren) are significant, as shown in Table A1 in the Appendix. The comparison of each

ABI child to the children with CI and NH indicates that children with CI and children with NH show a smaller preference for one place of articulation, in contrast to ABI1 and ABI2 who have a stronger preference for the coronal place of articulation, and ABI3 labial place of articulation in their target words.

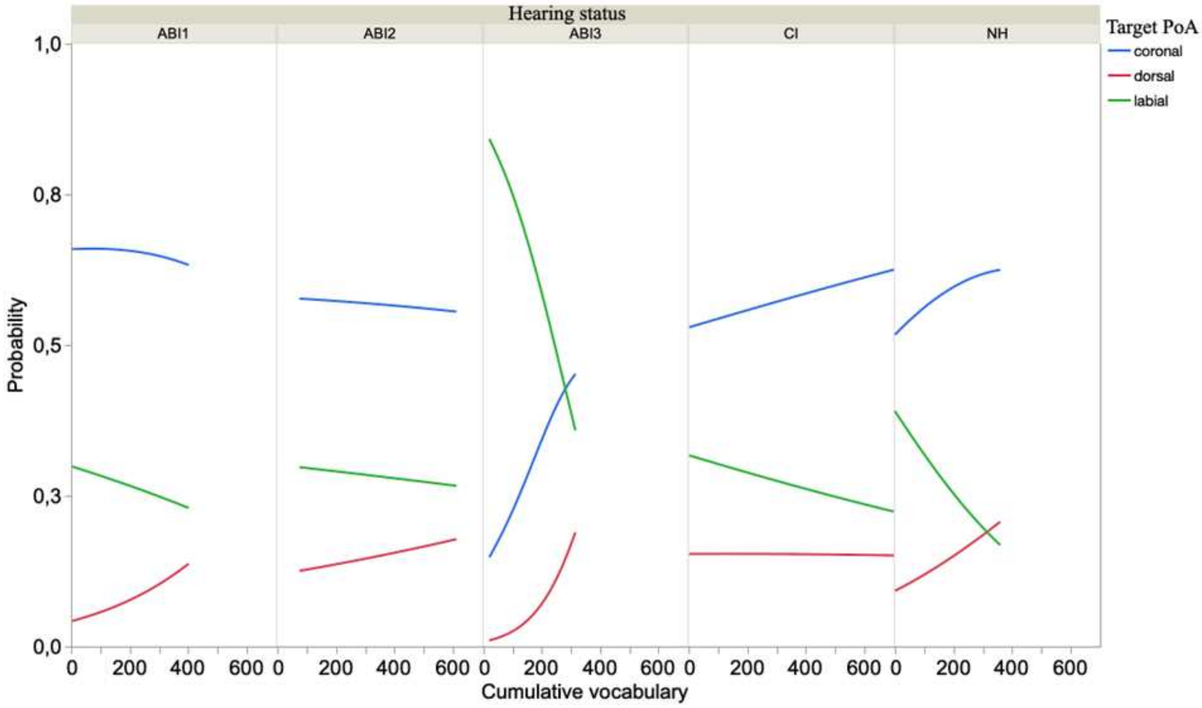


Figure 1: Frequency distribution in the targets - predicted probabilities

In children's own productions, there were four options in terms of place of articulation: labial, coronal, dorsal and omitted. In omission, no consonants were produced by the children, so the place of articulation could not be identified.

In all (groups of) children, the dorsal place of articulation is significantly less frequent than labial, coronal and omitted place of articulation (Table A2, Figure 2). For ABI1, ABI2 and the children with CI and NH, the coronal place of articulation is the most likely in children's own productions. The coronal place of articulation is significantly more frequent than labial and dorsal for ABI1 and ABI2. For children with CI and NH, these effects are significantly less pronounced, as can be seen in Figure 2 and the different main effects of Hearing status in Table A2 in the Appendix. When taking into account the omitted consonants as a place of articulation, ABI2 seems to follow the trend seen in children with CI and NH: coronal > labial > omitted > dorsal. The difference between the labial and omitted place of articulation is significant in ABI2, indicating that ABI2 was less likely to omit a consonant than to produce a labial consonant. This pattern is identical for the children with CI and NH, but it is significantly less pronounced (Table A2 in the Appendix). In contrast, ABI2 omitted consonants significantly more often than producing a labial consonant, whereas the opposite is true for children with CI and NH (Figure 2, Table A2).

As in the target, ABI3's most frequent place of articulation in the child's own productions is labial, which is significantly more frequent than the other places of articulation at a lexicon size of 100 word types (table A2). This effect decreases with increasing lexicon size.

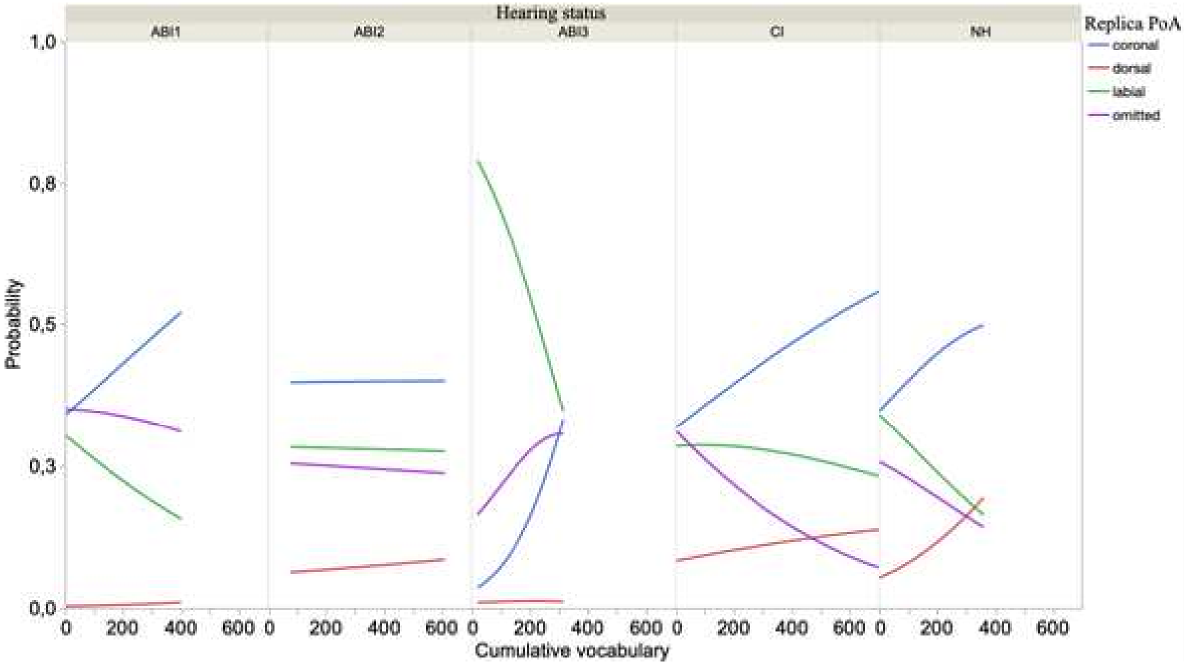


Figure 2: Frequency distribution in the replicas - predicted probabilities

Accuracy of place of articulation

The accuracy of the different places of articulation for each child with ABI and the control groups CI and NH are shown in Figure 3 and Table A3 in the Appendix. For all children with ABI, the labial place of articulation is produced significantly more accurately than the coronal and dorsal place of articulation, and the coronal place of articulation is produced significantly more accurately than the dorsal place of articulation. Overall, production accuracy increased significantly with increasing lexicon size, even though differences are represented in Figure 3 with respect to the different places of articulation. For each place of articulation, ABI2 is outperforming the other two ABI children and the difference becomes especially clear in the least accurately produced place of articulation, i.e., dorsal place of articulation. When comparing the children with ABI to one another in Figure 3, the difference in accuracy between the labial and coronal place of articulation is the least pronounced in ABI1 and the most pronounced in ABI3 (still, the differences remain significant for both children, see Table 2 in the Appendix). Yet, the difference between the labial and dorsal place of articulation as well as the coronal and dorsal place of articulation is the least outspoken in ABI2 (but still significant), whereas ABI1 and ABI3 show significant drops in accuracy rate for the dorsal place of articulation (Table A3 in the Appendix).

For children with CI and children with NH, the same trend is visible: the labial place of articulation is significantly more accurately produced than the coronal place of articulation, which is, in turn, significantly more accurately produced than the dorsal place of articulation. However, as can be derived from the interaction effects in Table A3 in the Appendix (PoA_Target x Hearing status) and as illustrated in Figure 3, the differences are significantly smaller in these children with CI and NH.

When comparing the children with ABI to the children with CI and NH, ABI children's production is significantly less accurate for all places of articulation. The difference with the children with CI and NH is smallest for the labial place of articulation, and most outspoken for the dorsal place of articulation. In this respect, ABI2 is most similar to CI and NH when compared to the other two ABI children.

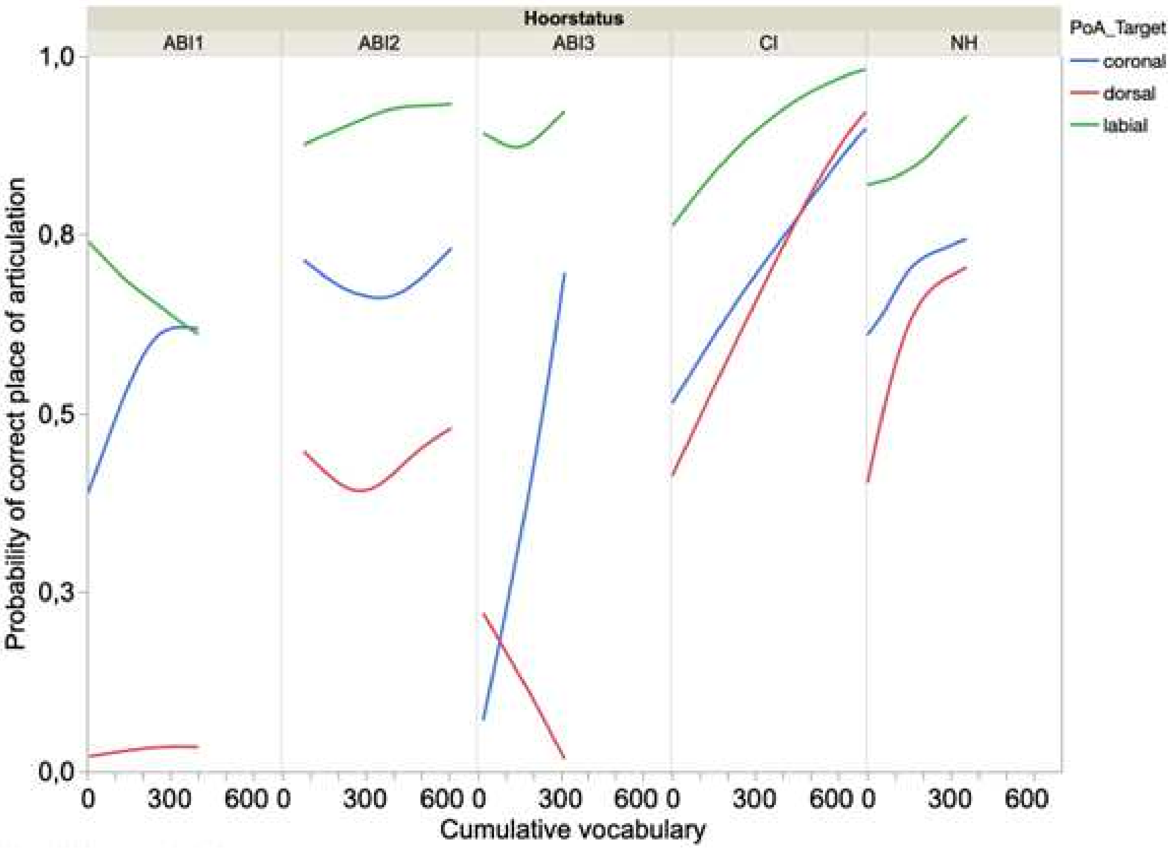


Figure 3: Probability of accurate production of place of articulation – observed values

Error analyses

Figure 4 represents the predicted error patterns for each ABI child and the control groups (CI and NH). For each target place of articulation produced incorrectly, the predicted development of erroneous place of articulation is displayed. The statistical analyses can be found in tables A4 – A6 in the Appendix. Overall, incorrect production of a place of articulation mostly involved omission of the consonant without replacing it by another place of articulation. This can be clearly seen in Figure 4 and is confirmed in the statistical analyses in Tables A4 - A6 in the Appendix.

For all ABI children as well as the children with CI and NH, an incorrectly produced labial consonant is more likely to be omitted than to be replaced by a coronal or a dorsal consonant (all significant, except for the coronal - omitted comparison in ABI1, Table A4 in the Appendix). ABI1 shows a somewhat different pattern in that coronal consonants appear more frequently to replace incorrect labials than they are omitted, even to a significantly greater extent than children with CI and children with NH.

Dorsal consonants appear very infrequently instead of labial ones in the error analyses. This difference is significant for all children (Table A4 in the Appendix), but the effect is significantly less pronounced in children with CI and children with NH (Table A4 in the Appendix) as compared to the children with ABI.

When comparing the ABI children to one another, Figure 4 indicates that ABI3 omitted the most frequently when labial place of articulation was produced incorrectly. ABI2, instead, seems to approach the trend in children with CI and NH the most. For instance, for the coronal - dorsal comparison in ABI2, there was no significant difference between ABI2 and the children with CI and NH (Table A4 in the Appendix).

When coronal place of articulation was not produced correctly, the consonant was significantly more often omitted than it was replaced by a labial or dorsal consonant (Table A5). Overall, the trend is highly similar in all children with ABI and children with CI and children with NH. However, statistical analyses showed that the effect is significantly less pronounced in children with CI and children with NH as compared to ABI1, ABI2 and ABI3 (please see Table A5 in the Appendix for the details).

When the consonant was replaced, it was significantly more often replaced by a labial consonant than by a dorsal one in ABI1, ABI2 and ABI3 (Table A5 in the Appendix). For ABI1, the effect is more pronounced than in children with CI (on the edge of significance, $p = 0.0487$) and children with NH (significant). For ABI2 and ABI3, the effect is more pronounced than in children with CI (significant), but less pronounced than in children with NH (only significant for ABI3). These differences can probably be explained by the higher omission rates in the children with ABI.

If a dorsal consonant was produced incorrectly in terms of place of articulation, it is significantly more likely that the consonant was omitted than it was replaced by a labial or a coronal consonant (Table A6 in the Appendix) in all children with ABI, CI and NH. Nevertheless, the effect is significantly less pronounced in children with CI and NH. To conclude, if the consonant was replaced, it was significantly more often replaced by a coronal consonant than a labial consonant for ABI1, ABI2 and the children with CI and NH. ABI3 showed a somewhat different pattern, in the sense that it was more likely that this child used a labial consonant than a coronal one in the very first word productions. Nevertheless, with increasing vocabulary size, this effect is reversed into the same trend as that was seen in all other children. At 100 word tokens, this is already significant (Table A6).

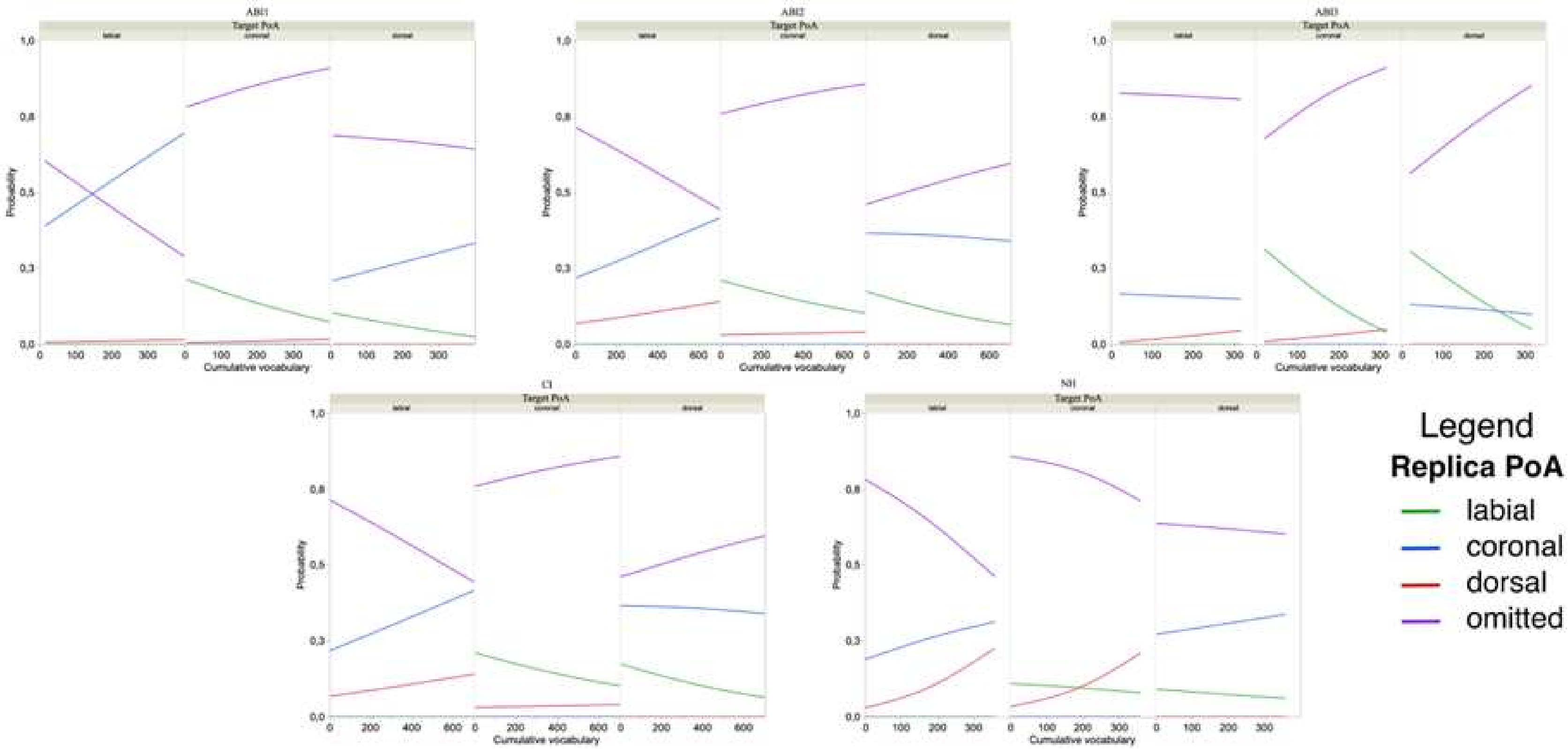


Figure 4: Predicted probability of the error patterns for all groups of children

Discussion

In this study, the production of place of articulation in spontaneous speech in three children with ABI were examined in comparison to two different control groups, i.e. children with CI and children with NH. This study used three general categories defining place of articulation: labial, coronal and dorsal. The place of articulation was identified for consonants in both the target word and the child's own production. Overall, the coronal place of articulation was the most frequently used in all children's (ABI, CI and NH) own productions as well as in the target words, except for ABI3, who showed a preference for labials (RQ1, RQ2). However, in terms of accuracy (RQ3), the labial place of articulation was more accurately produced than the coronal place of articulation by all children, and the dorsal place of articulation had a very low accuracy. In the error analyses (RQ4), it was shown that all places of articulation were most frequently omitted when they were produced incorrectly. If not omitted, the coronal place of articulation appeared as erroneous place of articulation.

Frequency of place of articulation in target and replica (RQ1 + RQ2)

With respect to children's target words and their own productions (replica), the results indicate that coronals were most frequent in all children, except for ABI3, who used more labials. The rank order between the three places of articulation was identical for ABI1, ABI2, children with CI and NH, i.e. coronals were more frequent than labials and dorsals. However, the differences are significantly smaller in children with CI and NH children in both targets and replicas. In other words, dorsals and labials appear more frequently in children with CI and children with NH than in ABI1 and ABI2. When comparing them to one another, ABI2 is approaching the probability levels of CI and NH more than ABI1.

These results complement the results of e.g. Beers (1995) for NH children, as well as Dillon, Cleary, et al. (2004); Dillon, Pisoni, et al. (2004) and Moreno-Torres and Moruno-Lopez (2014) for the CI – NH comparison: they found that coronals were more accurately produced than labials and dorsals in a non-word repetition task. In the present study, coronals were not produced more *accurately*, but more *frequently* and this was also clear for the children with CI and NH. This contrasts with the findings of Warner-Czyz and Davis (2008) who found that children with CI used more labials and children with NH used an equal proportion of labials and coronals.

The frequent occurrence of the coronal place of articulation in children's speech is not so surprising. In adult spoken Dutch, these sounds appear the most frequently as well (Luyckx et al., 2007). Thus, children hear more examples of coronal sounds, which could help with the acquisition. In addition, the labial and coronal place of articulation are generally acquired by children with NH before dorsal place of articulation (e.g. Beers, 1995). This high rate of coronal place of articulation fits also in the theory of Cummings et al. (2020). They stated, in line with the literature, that coronal place of articulation is the most frequent in languages, and that labial and dorsal place of articulation appear only with maturation and language exposure. In that sense, they suggested that children with phonological disorder, who are assumed to have less fine-grained phonological representations, are in an earlier stage of the developmental process, and have greater preference for coronals. In a similar vein, the preference for coronals and the smaller number of dorsals and labials in ABI children seems to suggest that they are in an earlier stage of speech development similar to children with phonological disorder in Cummings et al. (2020). ABI children have had less language exposure than the control groups. In addition, the speech signal provided by their implant is poor (Wong et al., 2019), so it is very likely that their phonological representations are less fine-grained as well. That would explain their delayed development.

Accuracy and error patterns (RQ3 + RQ4)

The coronal place of articulation appeared most frequently in children's spontaneous speech productions as well as their target words. Surprisingly, it was not the most accurately produced place of articulation. Instead, the labial place of articulation was significantly more often produced correctly than the coronal and dorsal place of articulation in all children. The effect is bigger in ABI children. When comparing the children with ABI to one another, the effect is the most pronounced in ABI1 and ABI3. ABI2 is approaching the levels of children with CI and NH to a greater extent, even though the difference is still significant. ABI1, on the other hand, showed the lowest accuracy rates overall, when taking all places of articulation into account. These findings, i.e. higher accuracy rates for labial consonants, are in line with those of Warner-Czyz and Davis (2008) with respect to children with CI as compared to children with NH: also in their study, labial consonants were produced more correctly by both groups of children as compared to other places of articulation.

This high likelihood of accurate labial production is inconsistent with Cummings et al. (2020). There are different ways to explain the high accuracy of labial sounds. In the first instance it is clear that labials are visually more salient than coronals and dorsals. As a result, it is likely that children learn to produce labials correctly more easily. As the effect is the most prominent in children with ABI, their exposure to sign language might be a contributing factor. Children with ABI are not only raised in spoken Dutch but receive a great amount of sign language input in order to communicate before and after implantation. Therefore, they have greater attention for cues this may result in more correct labials. This is supported by the fact that the effect is most prominent in ABI1 and ABI3, the two children with ABI most exposed to sign language. These children produce labials quite accurately, but have very low accuracy for the other two places of articulation, with a particularly low accuracy in dorsals.

Besides the factors mentioned above, it is possible that the inaccuracies in the production of labials and dorsals in children with ABI may have to be accounted for by a more limited set of word templates than children with CI and NH because of the later onset of hearing and the degraded speech signal. Vihman and Croft (2007) proposed a radical template phonology, in which they assume that children have so-called word templates, which are based on their own productions in the babbling period (vocal motor schemes) and the input they receive. In speech production, children adapt their word production to the word templates they have acquired. For instance, if a Dutch-speaking child has a word template /pV/, the child would produce /pu/ for /blum/ *bloem* (Eng. *flower*) and /pe/ for /spel/ *speel* (Eng. *to play*) and /ber/ *beer* (Eng. *teddy bear*) in order to produce the word in accordance to the acquired word template (Faes & Gillis, In press). This may well have an impact on the accuracy of place of articulation. Since ABI children have delayed and degraded auditory input as well as a less extensive babbling phase, it is likely that they have a smaller number of word templates. As a result, matching word productions to the few templates available, automatically increases the likelihood of inaccuracies in the production of consonants in general, as compared to children with CI and children with NH. With respect to children with ABI, the nature of their word templates remains an open question. Is it especially with labial consonants, given their high accuracy rate, or rather coronal consonants, given their frequency in target and replica (own production). In fact, the results of RQ1 and RQ2 and the results of RQ3 are somewhat contradictory. Yet, it is unclear how these differences are best explained, or which factors contribute to these effects. For instance, if it is sign language exposure, we would expect that labials appear more frequently in replicas. If it is a less developed phonology (in line with Cummings et al. (2020)), resulting in more use of the coronal place of articulation, coronal place of articulation would be expected to be pronounced the most accurately as well. It is open for future research to further disentangle these possibilities.

It should be pointed out that more accurate production of the labial place of articulation in all children was not reflected in the error patterns. Incorrectly produced consonants were most likely to be omitted by the ABI children. Although this effect was also noticeable in the control groups, it was significantly less outspoken. In other words, children with ABI and - to a lesser extent also - children with CI and NH, rather omit the consonant if they could not produce place of articulation correctly instead of producing it erroneously. This effect was also found in Warner-Czyz and Davis (2008) with respect to children with CI as compared to NH peers. In their study, children with CI omitted consonants with all places of articulation, whereas children with NH omitted especially labial and dorsal consonants, but no coronals – these were replaced. In other words, the CI group avoided complex phonology for a longer period than children with NH did. A similar finding was found for morphology: Szagun (2002) reported that CI children rather omit for instance articles and noun plurals instead of producing these morphological elements incorrectly. Children with NH, however, use nouns plurals and articles, but often incorrectly. Children with CI thus avoid using complex morphology initially, but they improve and catch up with children with NH with increasing hearing experience (Faes et al., 2015; Hammer, 2010). A similar trend seems to appear with the phonological development for especially children with ABI. They show an avoidance of complex phonology, mainly with respect to dorsal place of articulation, but also for labial and coronal place of articulation. So they rather omit a consonant than producing it incorrectly in terms of place of articulation. In this respect, it seems that children with ABI are avoiding complex phonology for an even longer time than children with CI (as compared to their NH peers in Warner-Czyz and Davis (2008)).

Individual variation

The children with ABI showed considerable individual variation, especially ABI3 showed a number of different patterns, while ABI1 and ABI2 were more similar. For instance, ABI3 used more labials than other places of articulation in both the target words as the replicas and that labials appear initially more often to replace erroneous dorsals. One of the factors that can contribute to this different pattern in ABI is probably to be found in the child's etiology. In contrast to ABI1 and ABI2, ABI3 wears both a CI and an ABI. In addition, ABI3 had different inner ear pathology leading to ABI implantation as compared to ABI1 and ABI2, which also resulted in later ABI implantation by the age of four after an unsuccessful CI period. CI implantation was no option for ABI1 and ABI2, so they received their ABI as early as their second birthday. It might be the case that these differences have led to another trend in ABI3's phonological development of place of articulation. In addition, the child received the ABI implant two years later than the other two children with ABI as the CI seemed not to have the desired effect, so the child and family relied on sign language much longer and more extensively. This might have caused the child's sensitivity for visual cues in the oral speech signal, leading to a general preference for the most salient place of articulation, viz. labial.

The differences with children with CI and children with NH remained obvious and significant even with increasing vocabulary size. Regarding the children with ABI, ABI2 was approaching the children with CI and the children with NH the most, even though these differences were also mostly significant. This is entirely in line with other research on the same children with ABI, CI and NH, in which ABI2 was also approaching the control groups with respect to for instance intraword variability (Faes & Gillis, In press), speech intelligibly (Faes, De Maeyer, et al., 2022), and phonemic accuracy (Faes & Gillis, 2021b). Yet, as stated in these studies, it is still unclear which factors are contributing to the better performance of ABI2 as compared to ABI1 overall. In the literature, factors such as hearing experience, age at implantation, absence of additional disabilities and lower hearing thresholds after implantation have been identified to explain individual differences between children with ABI. However, all these factors are identical in ABI1 and ABI2: they were both implanted by their second birthday and thus have a similar amount of hearing experience, none of them have additional disabilities and their hearing thresholds are also quite similar after implantation. So, further research on factors contributing to successful ABI use is desireable. This would be extremely useful in speech and language therapy applications for children with ABI (Hammes Ganguly et al., 2019).

Although place of articulation production of ABI2 is similar the children with CI and NH and shows similar trends, some caution should be taken with these results. In this study, lexical age has been used as a comparative measure between children with ABI and the controls. Although this seems a highly reasonable approach (viz., the close relationship between lexical and phonological development), this measure is not ideal either. The matching on lexicon size might have been the most optimistic measure, because when considering the hearing ages or chronological ages of the children, the differences with between children with ABI and children with NH and CI would increase even more. For instance, the children with NH in this study were not older than two years of age and were matched with children with ABI between three and six with similar lexicon sizes, which is considerably older in terms of chronological age. It seems very unlikely that a three-to-six-year-old child with NH performs in a similar vein as the current results for children with ABI. The same holds for children with CI: they were somewhat older in terms of chronological age as compared to the children with NH to reach the same lexicon sizes, but they were still younger than children with ABI at that point. In addition, they have had less hearing experience as well (fewer length of device use) (Faes & Gillis, 2019b). So, even though ABI2 seem to approach the trends visible the children with CI and the children with NH, there is still an enormous difference with respect to the children's chronological and hearing age. For ABI1 and ABI3, these effects are even bigger.

Conclusion

This study has shown that most children, except one child with ABI, used the coronal place of articulation the most frequently. This child with ABI most frequently used labials. However, all children produced labial place of articulation more accurately than the coronal place of articulation. Thus there was a difference between frequency and accuracy. Overall, the children with ABI performed the worst in terms of accuracy rates. It can be concluded that even though the children with ABI clearly benefit from their implant to develop spontaneous speech. However, there is still a long way to go for these children: they lag considerably behind their peers with CI and NH in speech production and speech production accuracy. Given these poor accuracy rates, it is reasonable that these children are lowly intelligible as well, as these aspects are closely related to one another (see e.g. Ingram, 2002). Therefore, it seems highly recommended for children with ABI and their environment to not only rely on oral communication. A valid option is for instance sign language use to guarantee efficient communication in this population, but choices need to be made according to the individual child's needs.

In-depth research disentangling the contributing factors for successful ABI use is desireable, both on a linguistic level to guide speech and language therapy (Hammes Ganguly et al., 2019) as on a medical level – for instance the precise placement and understanding of the possible tonotopic organization of the cochlear nucleus of the brainstem

Acknowledgements

This work was supported by the Research Foundation in Flanders (FWO) [grant 12Q6318N]. We would like to thank the children and their parents for participating in this study, and N. Boonen, I. Molemans, K. Schauwers, R. Van den Berg and L. Van Severen for the collection of most of the video-recordings.

References

Altvater-Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua, 120*, 1898-1909. https://doi.org/10.1016/j.lingua.2010.02.010

Aslan, F., Burcu Ozkan, H., Yücel, E., Sennaroglu, G., Bilginer, B., & Sennaroglu, L. (2020). Effects of age at auditory brainstem implantation: impact on auditory perception, language development, speech intelligibility. *Otology & Neurotology, 41*(1), 11 - 20. https://doi.org/10.1097/MAO.0000000000002455

Batuk, O. M., Cinar, B. C., Yarali, M., Aslan, F., Ozkan, H. B., Sennaroglu, G., Yucel, E., Bajin, M. D., Bilginer, B., & Sennaroglu, L. (2020). Bimodal stimulation in children with inner ear malformation: One side cochlear implant and contralateral auditory brainstem implant. *Clinical Otolaryngology, 45*, 231 - 238. https://doi.org/10.1111/coa.13499

Bayazit, Y., Kosaner, J., Cicek Cinar, B., Atac, A., Tutar, H., Gunduz, B., Altinyay, S., Gokdogan, C., Ant, A., Ozdek, A., & Goksu, N. (2014). Methods and preliminary outcomes of pediatric auditory brainstem implantation. *Annals of Otology, Rhinology & Laryngology, 123*(8), 529 - 536. https://doi.org/10.1177/0003489414525123

Beers, M. (1995). *The phonology of normally developing and language-impaired children Amsterdam University].* Unpublished doctoral dissertation.

Bergeson, T., Pisoni, B., & Davis, R. (2003). A longitudinal study of audivisual speech perception by children with hearing loss who have cochlear implants. *The Volta Review, 103*(4), 347-370.

Bergeson, T., Pisoni, B., & Davis, R. (2005). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. *Ear & Hearing, 26*(2), 149-164. https://doi.org/10.1097/00003446-200504000-00004

Blamey, P., Barry, J., Bow, C., Sarant, J., Paatsch, L., & Wales, R. (2001). The development of speech production following cochlear implantation. *Clinical Linguistics & Phonetics, 15*(5), 363 - 382. https://doi.org/10.1080/02699200010017823

Blamey, P., Barry, J., & Jacq, P. (2001). Phonetic inventory development in young cochlear implant users 6 years postoperation. *Journal of Speech Language and Hearing Research, 44*, 73 - 79. https://doi.org/10.1044/1092-4388(2001/007)

Booij, G. (1995). *The phonology of Dutch.* Clarendon Press.

Buchman, C. A., Teagle, H. F. B., Roush, P. A., Park, L. R., Hatch, D., Woodard, J., Zdanski, C., & Adunka, O. F. (2011). Cochlear implantation in children with labyrinthine anomalies and cochlear nerve deficiency: Implications for auditory brainstem implantation. *The Laryngoscope, 121*, 1979 - 1988. https://doi.org/10.1002/lary.22032

Colletti, L., Shannon, R., & Colletti, V. (2014). The development of auditory perception in children following auditory brainstem implantation. *Audiology and Neurotology, 19*(6), 386 - 394. https://doi.org/10.1159/000363684

Colletti, V., Fiorino, F., Carner, M., Miorelli, V., Guida, M., & Colletti, L. (2004). Perceptual outcomes in children with auditory brainstem implants. *International congress Series, 1273*, 425 - 428. https://doi.org/10.1016/j.ics.2004.07.047

Colletti, V., Fiorino, F., Sacchetto, L., Miorelli, V., & Carner, M. (2001). Hearing habilitation with auditory brainstem implantation in two children with cochlear nerve aplasia. *International Journal of Pediatric Otorhinolaryngology, 60*, 99 - 111. https://doi.org/10.1016/S0165-5876(01)00465-7

Cummings, A., Ogiela, D., & Wu Ying, C. (2020). Evidence for [coronal] underspecification in typical and atypical phonological development. *Frontiers in Human Neuroscience,* *14*. https://doi.org/10.3389/fnhum.2020.580697

Desai, S., Stickney, G., & Zeng, F. G. (2008). Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *The Journal of the Acoustical Society of America, 123*(1), 428 - 440. https://doi.org/https://doi.org/10.1121/1.2816573

Dillon, C., Cleary, M., Pisoni, D., & Carter, A. (2004). Imitation of nonwords by hearing-impaired children with cochlear implants: segmental analyses. *Clinical Linguistics & Phonetics, 18*(1), 39 - 55. https://doi.org/10.1080/0269920031000151669

Dillon, C., Pisoni, D., Cleary, M., & Carter, A. (2004). Nonword imitation by children with cochlear implants. *Archives of Otolaryngology–Head & Neck Surgery, 130*, 587-591. https://doi.org/10.1001/archotol.130.5.587

Drennan, W. R., & Rubinstein, J. T. (2008). Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of rehabilitation research and development, 45*(5), 779 - 790. https://doi.org/10.1682/JRRD.2007.08.0118

Eisenberg, L., Hammes Gangly, D., Martinez, A., Fisher, J. M., Winter, M., Glater, J., Schrader, D. K., Loggins, J., Wilkinson, E., & The Los Angeles Pediatric ABI Team. (2018). Early communication development of children with auditory brainstem implants. *Journal of Deaf Studies and Deaf Education, 23*(3), 249 – 260. https://doi.org/10.1093/deafed/eny010

Eriks-Brophy, A., Gibson, S., & Tucker, S. (2013). Articulatory error patterns and phonological process use of preschool children with and without hearing loss. *The Volta Review, 113*(2), 87-125.

Faes, J. (2017). *Speech production and speech production accuracy in young children: hearing and hearing impaired children with a cochlear implant* [PhD, University of Antwerp]. Antwerp.

Faes, J., De Maeyer, S., & Gillis, S. (2022). Speech intelligibility of children with an auditory brainstem implant: a triple-case study. *Clinical Linguistics & Phonetics.* https://doi.org/10.1080/02699206.2021.1988148

Faes, J., Gillis, J., & Gillis, S. (2015). Syntagmatic and paradigmatic development of cochlear implanted children in comparison with normally hearing peers up to age seven. *International Journal of Pediatric Otorhinolaryngology, 79*(9), 1533 - 1540. https://doi.org/10.1016/j.ijporl.2015.07.005

Faes, J., Gillis, J., & Gillis, S. (2022). Speech production accuracy of children with auditory brainstem implants: A comparison with peers with cochlear implants and typical hearing using Levenshtein distance. *First language, 42*, 22 - 50. https://doi.org/10.1177/01427237211042216

Faes, J., & Gillis, S. (2016). Word initial fricative production in children with cochlear implants and their normally hearing peers matched on lexicon size. *Clinical Linguistics & Phonetics, 30*(12), 959 - 982. https://doi.org/10.1080/02699206.2016.1213882

Faes, J., & Gillis, S. (2018). Language production outcomes after pediatric auditory brainstem implantation. *Journal of Hearing Science: Abstracts from the 15th International Conference on Cochlear Implants and Other Implantable Auditory Technologies*, Antwerp, 27 - 30 June 2018, 8(2), 374.

Faes, J., & Gillis, S. (2019a). Auditory brainstem implantation in children: effect on speech production. *International Journal of Pediatric Otorhinolaryngology, 119*, 103 - 112. https://doi.org/10.1016/j.ijporl.2019.01.014

Faes, J., & Gillis, S. (2019b). Expressive vocabulary growth after pediatric auditory brainstem implantation in two cases' spontaneous productions: a comparison with children with cochlear implants and typical hearing. *Frontiers in Pediatrics.* https://doi.org/10.3389/fped.2019.00191

Faes, J., & Gillis, S. (2021a). Consonant and vowel production in the spontaneous speech productions of children with auditory brainstem implants. *Clinical Linguistics & Phonetics, 35*(12), 1132 - 1160. https://doi.org/10.1080/02699206.2020.1869833

Faes, J., & Gillis, S. (2021b). Word characteristics and speech production accuracy in children with auditory brainstem implants: a longitudinal triple case report. *Clinical Linguistics & Phonetics, 35*(9), 874 - 890. https://doi.org/10.1080/02699206.2020.1838613

Faes, J., & Gillis, S. (In press). Intraword variability in children with auditory brainstem implants: a longitudinal comparison with children with typical hearing and cochlear implants. *American Journal of Speech-Language Pathology.*

Farhood, Z., Nguyen, S. A., Miller, S. C., Holcomb, M. A., Meyer, T. A., & Rizk, H. G. (2017). Cochlear implantation in inner ear malformations: systematic review of speech perception outcomes and intraoperative findings. *Otolaryngology - Head and Neck Surgery, 156*, 783 - 793. https://doi.org/10.1177/0194599817696502

Friedman, D. R., Asfour, L., Shapiro, W., Thomas Roland, J., & Waltzman, S. (2018). Performance with an auditory brainstem implant and contralateral cochlear implant in pediatric patients. *Audiology and Neurotology, 23*, 216 - 221. https://doi.org/10.1159/000493085

Hammer, A. (2010). *The acquisition of verbal morphology in cochlear-implanted and specific language impaired children* [University of Leiden]. Unpublished doctoral dissertation.

Hammes Ganguly, D., Schrader, D. K., & Martinez, A. (2019). Planning for and working with children with an auditory brainstem implant: what therapists need to know. *Perspectives of the ASHA Special Interest Groups, 4*, 149 - 166. https://doi.org/10.1044/2018_-PERS-SIG-2018-0002

Huyse, A., Berthommier, F., & Leybaert, J. (2013). Degradation of labial information modifies audivisual speech perception in cochlear-implanted children. *Ear & Hearing, 34*(1), 110-121. https://doi.org/10.1097/AUD.0b013e3182670993

Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language, 29*, 713-733. https://doi.org/10.1017/S0305000902005275

Jusczyk, P. (1992). Developing phonological categories from the speech signal. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological Development: Models, Research and Implications.* York Press Timonium.

Long, C., Nimmo-Smoth, I., Baguley, D., O'Driscoll, M., Ramsden, R., Otto, S., Axon, P., & Carlyon, R. (2005). Optimizing the clinical fit of auditory brain stem implants. *Ear & Hearing, 26*(3), 251 - 262. https://doi.org/10.1097/00003446-200506000-00002

Luyckx, K., Kloots, H., Coussé, E., & GIllis, S. (2007). Klankfrequenties in het Nederlands. In D. Sandra, R. Rymenans, P. Cuvelier, & P. Van Petegem (Eds.), *Tussen taal, spelling en* *onderwijs* (pp. 141 - 154). Academia Press.

MacWhinney, B. (2000). T*he CHILDES project: tools for analyzing talk.* NJ: Lawrence Erlbaum Associates.

Mertens, P. (2001). *Fonilex*. Retrieved 14 august from http://bach.arts.kuleuven.be/fonilex/

Molemans, I. (2011). *A longitudinal investigation of aspects of the prelexical speech repertoire in young children acquiring Dutch: normally hearing children and hearing-impaired children with a cochlear implant* [PhD, University of Antwerp]. Antwerp.

Moreno-Torres, I., & Moruno-Lopez, E. (2014). Segmental and suprasegmental errors in Spanish learning cochlear implant users: neurolinguistic interpretation. *Journal of Neurolinguistics, 31*, 1 - 16. https://doi.org/10.1016/j.jneuroling.2014.04.002

Moreno-Torres, I., & Moruno-López, E. (2014). Segmental and suprasegmental errors in Spanish learning cochlear implant users: Neurolinguistic interpretation. *Journal of Neurolinguistics, 31*, 1-16. https://doi.org/10.1016/j.jneuroling.2014.04.002

Most, T., Rothem, H., & Luntz, M. (2009). Auditory, visual, and auditory-visual speech perception by individuals with cochlear implants versus individuals with hearing aids. *American Annals of the Deaf, 154*(3), 284 - 292. https://doi.org/10.1353/aad.0.0098

Schauwers, K. (2006). *Early speech and language development in deaf children with a cochlear implant: a longitudinal investigation* [PhD, University of Antwerp]. Antwerp.

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the national academy of sciences, 102*(51), 18748-18750. https://doi.org/10.1073/pnas.0508862102

Sennaroglu, L., Colletti, V., Lenarz, T., Manrique, M., Laszig, R., Rask-Andersen, H., Gösku, N., Offeciers, E., Saeed, S., Behr, R., Bayazit, Y., Casselman, J., Freeman, S., Kileny, P., Lee, D., Shannon, R., Kameswaran, M., Hagr, A., Zarowski, A., Schwartz, M., Bilginer, B., Kishore, A., Sennaroglu, G., Yücel, E., Sarac, S., Atas, A., Colletti, L., O'Driscoll, M., Moon, I., Gärtner, L., Huarte, A., Nyberg, G., Özgen Mocan, B., Atay, G., Demir Bajin, M., Cicek Cinar, B., Özbal Batuk, M., Yarali, M., Aydinli, F., Aslan, F., Kirazli, M., Özkan, B., Hans, J., Kosaner, J., & Polak, M. (2016). Consensus statement: long-term results of ABI in children with complex inner ear malformations and decision making between CI and ABI. *Cochlear Implants International, 17*(4), 163 - 171. https://doi.org/10.1080/14670100.2016.1208396

Sennaroglu, L., Sennaroglu, G., Yücel, E., Bilginer, B., Atay, G., Demir Bajin, M., Özgen Mocan, B., Yarali, M., Aslan, F., Cicek Cinar, B., Özkan, B., Özbal Batuk, M., Ekin Kirazli, C., Karakaya, J., Atas, A., Sarac, S., & Ziyal, I. (2016). Long-term results of ABI in children with severe inner ear malformations. *Otology & Neurotology, 37*, 865 - 872. https://doi.org/10.1097/MAO.0000000000001050

Smith, B., McGregor, K., & Demille, D. (2006). Phonological development in lexically precocious 2-year-olds. *Applied Psycholinguistics, 27*, 355 - 375. https://doi.org/10.1017/S0142716406060310

Sosa, A., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language and Hearing Research, 55*, 596 - 608. https://doi.org/10.1044/1092-4388(2011/10-0113)

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language, 38*, 1-34. https://doi.org/10.1017/S0305000910000425

Stoel-Gammon, C., & Sosa, A. (2007). Phonological development. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 238-256). Blackwell Publishing.

Szagun, G. (2002). The acquisition of grammar in young German-speaking children with cochlear implants and with normal hearing. *Antwerp papers in linguistics, 102*, 40 - 60.

Szagun, G., & Stumper, B. (2012). Age or experience? The influence of age at implantation and social and linguistic environment on language development in children with cochlear implants. *Journal of Speech, Language and Hearing Research, 55*, 1640 - 1654. https://doi.org/10.1044/1092-4388(2012/11-0119)

Teagle, H. F. B., Henderson, L., He, S., Ewend, M. G., & Buchman, C. A. (2018). Pediatric auditory brainstem implantation: surgical, electrophysiologic, and behavioral outcomes. *Ear & Hearing, 39*, 326 - 336. https://doi.org/10.1097/AUD.0000000000000501

van den Berg, R. (2012). *Syllables inside out: a longitudinal study of the development of syllable types in toddlers acquiring Dutch: a comparison between hearing impaired children with a cochlear implant and normally hearing children* [PhD, University of Antwerp]. Antwerp.

van der Straaten, T., Netten, A., Boermans, P., Briaire, J., Scholing, E., Koot, R., Malessy, M., van der Mey, A., Verbist, B., & Frijns, J. (2019). Pediatric auditory brainstem implant users compared with cochlear implant users with additional disabilities. *Otology & Neurotology, 40*(7), 936 - 945. https://doi.org/10.1097/MAO.0000000000002306

Van Severen, L. (2012). *A large-scale longitudinal survey of consonant development in toddlers' spontaneous speech* [PhD, University of Antwerp]. Antwerp.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S.* (4th ed.). Springer. https://doi.org/10.1007/b97626

Vihman, M., & Croft, W. (2007). Phonological development: toward a 'radical' templatic phonology. *Linguistics, 45*, 683-725. https://doi.org/10.1515/LING.2007.021

Warner-Czyz, A., & Davis, B. (2008). The emergence of segmental accuracy in young cochlear implant recipients. *Cochlear Implants International, 9*(3), 143 - 166. https://doi.org/10.1002/cii.364

Wong, K., Kozin, E., Kanumuri, V., Vachicouras, N., Miller, J., Lacour, S., Brown, C., & Lee, D. (2019). Auditory brainstem implants: recent progress and future perspectives. *Frontiers in Neuroscience, 13*:10. https://doi.org/10.3389/fnins.2019.00010

Yucel, E., Aslan, F., Ozkan, H., & Sennaroglu, L. (2015). Recent rehabilitation experience with pediatric ABI users. *The Journal of International Advanced Otology, 11*, 110 - 113. https://doi.org/10.5152/iao.2015.915

Appendix

Table A1: Statistical analyses for target place of articulation: likelihood of labial place of articulation (vs. coronal, vs. dorsal) and coronal place of articulation (vs. dorsal)

	ABI1			**ABI2**			**ABI2**		
Labial vs. Coronal	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*
Intercept	0.85	0.06	<0.0001	0.42	0.01	<0.0001	-0.05	0.01	<0.0001
Cumulative vocabulary	<0.01	<0.01	0.1130	<0.01	<0.01	<0.0001	<0.01	<0.01	<0.0001
Hearing status CI	-0.43	0.07	<0.0001	0.04	0.01	0.0103	0.47	0.02	<0.0001
Hearing status NH	-0.29	0.06	<0.0001	0.14	0.01	<0.0001	0.62	0.02	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	<0.0001	<0.01	<0.01	<0.0001	<-0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<0.01	<0.01	<0.0001	<0.01	<0.01	<0.0001	<0.01	<0.01	0.4928
Labial vs. Dorsal									
Intercept	-1.61	0.13	<0.0001	-0.46	0.01	<0.0001	-1.29	0.03	<0.0001
Cumulative vocabulary	<0.01	<0.01	<0.0001	<0.01	<0.01	<0.0001	<0.01	<0.01	<0.0001
Hearing status CI	0.82	0.13	<0.0001	-0.13	0.02	<0.0001	0.48	0.04	<0.0001
Hearing status NH	0.62	0.13	<0.0001	-0.33	0.02	<0.0001	0.30	0.04	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	<0.0001	<0.01	<0.01	<0.0001	<-0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<0.01	<0.01	0.1344	<0.01	<0.01	<0.0001	<0.01	<0.01	0.4592
Coronal vs. Dorsal									
Intercept	-2.46	0.12	<0.0001	-1.07	0.01	<0.0001	-1.24	0.03	<0.0001
Cumulative vocabulary	<0.01	<0.01	<0.0001	<0.01	<0.01	<0.0001	<-0.01	<0.01	<0.0001
Hearing status CI	1.25	0.12	<0.0001	-0.17	0.02	<0.0001	<0.01	0.04	0.8700
Hearing status NH	0.91	0.12	<0.0001	-0.48	0.02	<0.0001	-0.31	0.04	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	<0.0001	<-0.01	<0.01	<0.0001	<-0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<-0.01	<0.01	0.0107	<0.01	<0.01	<0.0001	<0.01	<0.01	0.7172

Cumulative vocabulary is centered at 100 word tokens

Table A2: Statistical analyses for replica place of articulation: likelihood of labial place of articulation (vs. coronal, vs. dorsal, vs. omitted), coronal place of articulation (vs. dorsal, vs. omitted) and dorsal place of articulation (vs. omitted)

	ABI1			**ABI2**			**ABI2**		
Labial vs. Coronal	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*
Intercept	0.38	0.07	<0.0001	0.18	0.01	<0.001	-0.47	0.02	<0.0001
Cumulative vocabulary	<0.01	<0.01	<0.0001	<0.01	<0.01	<0.001	<0.01	<0.01	<0.0001
Hearing status CI	-0.34	0.07	<0.0001	-0.11	<0.01	<0.001	0.51	0.03	<0.0001
Hearing status NH	-0.06	0.07	0.4334	0.14	<0.01	<0.001	0.80	0.02	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.0380	<0.01	<0.01	<0.001	<-0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<0.01	<0.01	0.4048	<0.01	<0.01	<0.001	<-0.01	<0.01	<0.0001
Labial vs. Dorsal									
Intercept	-4.07	0.01	<0.0001	-1.02	0.02	<0.001	-1.68	0.05	<0.0001
Cumulative vocabulary	<0.01	<0.01	<0.0001	<0.01	<0.01	<0.001	<0.01	<0.01	<0.0001
Hearing status CI	2.75	0.02	<0.0001	-0.30	0.03	<0.001	0.35	0.06	<0.0001
Hearing status NH	2.78	0.02	<0.0001	-0.26	0.02	<0.001	0.39	0.05	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.0008	<-0.01	<0.01	0.0709	<-0.01	<0.01	0.0458
Cumulative vocabulary * Hearing status NH	<0.01	<0.01	0.1386	<0.01	<0.01	<0.001	<0.01	<0.01	<0.0001
Coronal vs. Dorsal									
Intercept	-4.45	0.01	<0.0001	-1.21	0.02	<0.001	-1.21	0.06	<0.0001
Cumulative vocabulary	<0.01	<0.01	0.0291	<0.01	<0.01	<0.001	<-0.01	<0.01	0.0049
Hearing status CI	3.09	0.02	<0.0001	-0.19	0.02	<0.001	-0.16	0.06	0.0073
Hearing status NH	2.84	0.02	<0.0001	-0.40	0.02	<0.001	-0.41	0.06	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.0155	<-0.01	<0.01	<0.001	<0.01	<0.01	0.0146
Cumulative vocabulary * Hearing status NH	<0.01	<0.01	0.2747	<0.01	<0.01	<0.001	<0.01	<0.01	<0.0001
Labial vs. Omitted									
Intercept	0.28	0.07	<0.0001	-0.15	0.01	<0.001	-0.41	0.01	<0.0001
Cumulative vocabulary	<0.01	<0.01	0.0004	<-0.01	<0.01	0.0016	<0.01	<0.01	<0.0001
Hearing status CI	-0.52	0.08	<0.0001	-0.09	0.02	<0.001	0.19	0.02	<0.0001
Hearing status NH	-0.52	0.07	<0.0001	-0.10	0.01	<0.001	0.17	0.02	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	<0.0001	<-0.01	<0.01	<0.001	<-0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<-0.01	<0.01	0.0172	<0.01	<0.01	<0.001	<-0.01	<0.01	0.0005
Coronal vs. Omitted									
Intercept -0.10	0.07	0.1134	-0.33	0.01	<0.001	0.07	0.02	0.0056	
cumulative vocabulary	<-0.01	<0.01	<0.0001	<-0.01	<0.01	<0.001	<-0.01	<0.01	<0.0001
Hearing status CI	-0.19	0.07	0.0065	0.02	0.02	0.1392	-0.32	0.03	<0.0001
Hearing status NH	-0.46	0.07	<0.0001	-0.23	0.01	<0.001	-0.63	0.03	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.0001	<-0.01	<0.01	<0.001	<-0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<-0.01	<0.01	0.0002	<-0.01	<0.01	<0.001	<0.01	<0.01	0.1458
Dorsal vs. Omitted									
Intercept	4.34	0.04	<0.0001	0.88	0.02	<0.001	1.27	0.05	<0.0001
Cumulative vocabulary	<-0.01	<0.01	0.0001	<-0.01	<0.01	<0.001	<-0.01	<0.01	<0.0001
Hearing status CI	-3.28	0.05	<0.0001	0.21	0.03	<0.001	-0.16	0.06	0.0055
Hearing status NH	-3.30	0.04	<0.0001	0.17	0.02	<0.001	-0.23	0.06	<0.0001
Cumulative vocabulary * Hearing status CI	<0.01	<0.01	0.4697	<-0.01	<0.01	<0.001	<-0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<-0.01	<0.01	0.0088	<-0.01	<0.01	<0.001	<-0.01	<0.01	<0.0001

Cumulative vocabulary is centered at 100 word tokens

Table A3: Statistical analyses of production accuracy

	ABI1			**ABI2**			**ABI2**		
Labial vs. Coronal, Labial vs. Dorsal	*Estimate*	*std. Error*	*p*	*Estimate*	*std. Error*	*p*	*Estimate*	*std. Error*	*p*
Intercept (labial)	0.22	0.06	<0.0001	1.76	0.05	<0.001	1.73	0.06	<0.0001
Cumulative vocabulary	<0.01	<0.01	<0.0001	<0.01	<0.01	<0.001	<0.01	<0.01	<0.0001
Hearing status CI	1.32	0.07	<0.0001	-0.15	0.05	0.0045	-0.23	0.06	0.0003
Hearing status NH	1.34	0.07	<0.0001	-0.20	0.05	0.0001	-0.17	0.06	0.0072
PoA_Target coronal	-0.28	0.07	<0.0001	-1.66	0.05	<0.001	-2.12	0.08	<0.0001
PoA_Target dorsal	-4.14	0.27	<0.0001	-2.81	0.06	<0.001	-5.20	0.22	<0.0001
PoA_Target coronal * Hearing status CI	-0.95	0.08	<0.0001	0.37	0.06	<0.001	0.93	0.08	<0.0001
PoA_Target dorsal * Hearing status CI	2.63	0.27	<0.0001	1.37	0.07	<0.001	3.67	0.23	<0.0001
PoA_Target coronal * Hearing status NH	-0.63	0.08	<0.0001	0.74	0.05	<0.001	1.20	0.08	<0.0001
PoA_Target dorsal * Hearing status NH	2.82	0.27	<0.0001	1.49	0.07	<0.001	3.85	0.22	<0.0001
Coronal vs. Dorsal									
Intercept (coronal)	-0.06	0.04	<0.0001	0.11	0.02	<0.001	-0.39	0.05	<0.0001
Cumulative vocabulary	<0.01	<0.01	<0.0001	<0.01	<0.01	<0.001	<0.01	<0.01	<0.0001
Hearing status CI	0.36	0.04	<0.0001	0.22	0.02	<0.001	0.70	0.05	<0.0001
Hearing status NH	0.71	0.04	<0.0001	0.56	0.02	<0.001	1.04	0.05	<0.0001
PoA_Target dorsal	-3.86	0.27	<0.0001	-1.13	0.04	<0.001	-3.08	0.22	<0.0001
PoA_Target dorsal * Hearing status CI	3.58	0.27	<0.0001	1.00	0.05	<0.001	2.74	0.22	<0.0001
PoA_Target dorsal * Hearing status NH	3.45	0.27	<0.0001	0.73	0.05	<0.001	2.65	0.22	<0.0001

Cumulative vocabulary is centered at 100 word tokens

Table A4: Error patterns when labial place of articulation was produced incorrectly

	ABI1			**ABI2**			**ABI2**		
Coronal vs. Dorsal	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*
Intercept	-13.23	0.5	<0.0001	-1.05	0.12	<0.001	-1.02	0.16	<0.0001
Cumulative vocabulary	0.03	<0.01	<0.0001	<0.01	<0.01	0.0780	<-0.01	<0.01	0.9624
Hearing status CI	11.80	0.10	<0.0001	-0.18	0.15	0.2149	-0.45	0.19	0.0161
Hearing status NH	11.97	0.08	<0.0001	-0.21	0.14	0.1232	-0.24	0.17	0.1604
Cumulative vocabulary * Hearing status CI	-0.03	<0.01	<0.0001	<-0.01	<0.01	0.7562	<0.01	<0.01	0.0265
Cumulative vocabulary * Hearing status NH	-0.03	<0.01	<0.0001	<0.01	<0.01	0.1532	<0.01	<0.01	0.2927
Coronal vs. Omitted									
Intercept	0.17	0.20	0.4058	0.80	0.06	<0.001	1.18	0.07	<0.0001
Cumulative vocabulary	<-0.01	<0.01	<0.0001	<-0.01	<0.01	<0.001	<-0.01	<0.01	<0.0001
Hearing status CI	0.94	0.21	<0.0001	0.25	0.08	0.0011	-0.08	0.09	0.3379
Hearing status NH	1.00	0.21	<0.0001	0.37	0.07	<0.001	-0.02	0.08	0.8425
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.6499	<-0.01	<0.01	0.0589	<0.01	<0.01	0.1027
Cumulative vocabulary * Hearing status NH	<-0.01	<0.01	0.1285	<-0.01	<0.01	<0.001	<-0.01	<0.01	0.2659
Omitted vs. Dorsal									
Intercept	-13.40	0.05	<0.0001	-1.85	0.11	<0.001	-2.21	0.14	<0.0001
Cumulative vocabulary	0.04	<0.01	<0.0001	<0.01	<0.01	<0.001	<0.01	<0.01	<0.0001
Hearing status CI	10.86	0.09	<0.0001	-0.43	0.14	0.0020	-0.37	0.17	0.0333
Hearing status NH	10.97	0.07	<0.0001	-0.58	0.13	<0.001	-0.22	0.15	0.1515
Cumulative vocabulary * Hearing status CI	-0.03	<0.01	<0.0001	<0.01	<0.01	0.4194	<0.01	<0.01	0.1251
Cumulative vocabulary * Hearing status NH	-0.03	<0.01	<0.0001	<0.01	<0.01	<0.001	<0.01	<0.01	0.0938

Cumulative vocabulary is centered at 100 word tokens

Table A5: Error patterns when coronal place of articulation was produced incorrectly

	ABI1			**ABI2**			**ABI2**		
Labial vs. Dorsal	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*
Intercept	-2.64	0.46	<0.0001	-0.61	0.05	<0.001	-1.15	0.11	<0.0001
Cumulative vocabulary	<0.01	<0.01	0.2853	<0.01	<0.01	<0.001	<0.01	<0.01	<0.0001
Hearing status CI	0.93	0.47	0.0487	-1.21	0.08	<0.001	-0.52	0.13	<0.0001
Hearing status NH	2.10	0.46	<0.0001	0.06	0.06	0.3234	0.61	0.12	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.3644	<0.01	<0.01	0.0980	<-0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<0.01	<0.01	0.1415	<0.01	<0.01	<0.001	<0.01	<0.01	0.0394
Labial vs. Omitted									
Intercept	1.57	0.13	<0.0001	1.41	0.03	<0.001	1.24	0.04	<0.0001
Cumulative vocabulary	<0.01	<0.01	<0.0001	<0.01	<0.01	0.0583	<0.01	<0.01	0.0006
Hearing status CI	-0.21	0.14	0.1261	-0.06	0.04	0.1404	0.17	0.05	0.0004
Hearing status NH	0.53	0.13	<0.0001	0.69	0.04	<0.001	0.86	0.05	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.0320	<0.01	<0.01	<0.001	<-0.01	<0.01	0.0031
Cumulative vocabulary * Hearing status NH	<-0.01	<0.01	0.0017	<0.01	<0.01	0.6448	<-0.01	<0.01	0.0180
Omitted vs. Dorsal									
Intercept	-4.21	0.45	<0.0001	-2.02	0.04	<0.001	-2.39	0.11	<0.0001
Cumulative vocabulary	<-0.01	<0.01	0.8636	<0.01	<0.01	<0.001	<0.01	<0.01	0.0011
Hearing status CI	1.13	0.46	0.0132	-1.15	0.07	<0.001	-0.69	0.12	<0.0001
Hearing status NH	1.56	0.45	0.0005	-1.63	0.05	<0.001	-2.25	0.11	0.0266
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.8105	<-0.01	<0.01	0.0116	<-0.01	<0.01	0.0008
Cumulative vocabulary * Hearing status NH	<0.01	<0.01	0.0085	<0.01	<0.01	<0.001	<0.01	<0.01	0.0003

Cumulative vocabulary is centered at 100 word tokens

Table A6: Error patterns when dorsal place of articulation was produced incorrectly

	ABI1			**ABI2**			**ABI2**		
Labial vs. Coronal	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*	*coeff*	*stdv*	*p*
Intercept	-2.64	0.46	<0.0001	-0.61	0.05	<0.001	-1.15	0.11	<0.0001
Intercept	0.89	0.40	0.0269	1.18	0.04	<0.001	0.49	0.08	<0.0001
Cumulative vocabulary	<0.01	<0.01	0.0035	<0.01	<0.01	0.5475	<0.01	<0.01	0.0108
Hearing status CI	-0.06	0.41	0.8918	-0.32	0.06	<0.001	0.27	0.10	0.0003
Hearing status NH	0.42	0.41	0.3014	0.12	0.06	0.0482	0.83	0.10	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.0417	<0.01	<0.01	0.0070	<-0.01	<0.01	0.1793
Cumulative vocabulary * Hearing status NH	<-0.01	<0.01	0.0134	<0.01	<0.01	0.5067	<-0.01	<0.01	0.5245
Labial vs. Omitted									
Intercept	1.97	0.36	<0.0001	2.02	0.04	<0.001	0.77	0.08	<0.0001
Cumulative vocabulary	<0.01	<0.01	0.0188	<0.01	<0.01	0.2564	<0.01	<0.01	<0.0001
Hearing status CI	-0.72	0.37	0.0516	-0.83	0.06	<0.001	0.51	0.10	<0.0001
Hearing status NH	0.05	0.37	0.8866	<0.01	0.06	0.9444	1.26	0.10	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.0607	<0.01	<0.01	0.020	<-0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<-0.01	<0.01	0.1176	<0.01	<0.01	0.0868	<-0.01	<0.01	0.0002
Omitted vs. Coronal									
Intercept	-1.08	0.24	<0.0001	-0.84	0.05	<0.001	-0.29	0.08	<0.0001
Cumulative vocabulary	<0.01	<0.01	0.1563	<-0.01	<0.01	0.4630	<-0.01	<0.01	<0.0001
Hearing status CI	0.67	0.25	0.0076	0.51	0.06	<0.001	-0.14	0.09	0.1389
Hearing status NH	0.37	0.25	0.1346	0.12	0.06	0.0421	-0.43	0.09	<0.0001
Cumulative vocabulary * Hearing status CI	<-0.01	<0.01	0.5963	<-0.01	<0.01	0.6982	<0.01	<0.01	<0.0001
Cumulative vocabulary * Hearing status NH	<-0.01	<0.01	0.0505	<-0.01	<0.01	0.0991	<0.01	<0.01	<0.0001

Cumulative vocabulary is centered at 100 word tokens

STEM-, SPRAAK- EN TAALPATHOLOGIE
Vol. 27, 2022, Special Issue, pp. 177-205
Special Issue for Steven Gillis
Gepubliceerd online: 5 november 2022

University of Groningen Press
https://doi.org/10.21827/32.8310/2022-SG-177

Dutch Norming Study For 208 Color Drawings Depicting Transitive Events

Edwige Sijyeniyo[1], Robert J. Hartsuiker[2], & Sarah Bernolet[1]

[1] *Department of Linguistics, University of Antwerp*
[2] *Department of Experimental Psychology, Ghent University*

Samenvatting

We hebben een normeringsstudie uitgevoerd onder Nederlandse kinderen (6-12 jaar) om te onderzoeken of zij onze set van 208 kleurentekeningen, ontwikkeld voor longitudinaal onderzoek naar de productie en het begrip van transitieve zinnen, correct zouden interpreteren als de transitieve actie op elke tekening werd beschreven met een actieve of een passieve zin.

De kinderen gaven in 93.02% van de gevallen een correct antwoord voor onze tekeningen, wat aangeeft dat de tekeningen duidelijk zijn in termen van hoe de transitieve acties worden weergegeven. Er waren verschillende factoren die bijdroegen tot een incorrecte interpretatie van een overgankelijke actie in een plaatje. (1) Afbeeldingen waren moeilijker te interpreteren wanneer een passieve zin werd gebruikt om een overgankelijke gebeurtenis te beschrijven dan wanneer een actieve zin werd gebruikt; (2) kinderen maakten meer fouten bij het interpreteren van afbeeldingen wanneer abstracte werkwoorden, zoals "inhalen" en "vervangen", werden gebruikt om een overgankelijke gebeurtenis te beschrijven; (3) kinderen hadden moeite met het interpreteren van onze afbeeldingen wanneer bepaalde werkwoorden ("roepen" en "volgen") werden gepresenteerd in een passieve zin, terwijl andere werkwoorden ("blokkeren" en "inhalen") *meer* correcte antwoorden opleverden als de overgankelijke gebeurtenis werd beschreven met een passieve zin; ten slotte, (4) gaven jongere kinderen meer incorrecte antwoorden op onze afbeeldingen dan oudere kinderen, ongeacht of de overgankelijke handeling in een afbeelding werd beschreven met een actieve of een passieve zin. De jongere kinderen gaven vooral significant meer incorrecte antwoorden op plaatjes met het werkwoord "groeten" in vergelijking met de oudere kinderen.

Vanwege de vele werkwoord-naamwoord combinaties in onze set is deze erg nuttig voor onderzoekers die de verwerving van transitieve structuren bij kinderen willen onderzoeken, niet alleen voor studies met typisch ontwikkelende kinderen, maar ook voor studies met kinderen die een achterstand in hun taalontwikkeling hebben.
Trefwoorden: Normeringsstudie, set van kleurentekeningen, transitieve structuren, L1 syntaxverwerving

Correspondentieadres:
Edwige Sijyeniyo
Department of Linguistics
University of Antwerp
Prinsstraat 13, 2000 Antwerp
Belgium
E-mail: Edwige.Sijyeniyo@uantwerpen.be

Abstract

We conducted a norming study amongst Dutch children (6-12 years) to investigate whether they would correctly interpret our set of 208 color drawings, developed for longitudinal research on the production and comprehension of transitive sentences, if the transitive action on each drawing was described with an active or a passive sentence.

The children provided 93.02% correct answers to our pictures, which indicates that the pictures are clear in terms of how the transitive actions are displayed. There were several factors that contributed to an incorrect interpretation of a transitive action in a picture. (1) Pictures were more difficult to interpret when a passive sentence was used to describe a transitive event compared to an active sentence; (2) children made more errors interpreting pictures when abstract verbs, such as "inhalen" (to overtake) and "vervangen" (to replace), were used to describe a transitive event; (3) children had difficulty interpreting our pictures when particular verbs ("roepen" [to shout] and "volgen" [to follow]) were presented in a passive sentence, while other verbs ("blokkeren" [to block] and "inhalen") rendered more *correct* answers when the transitive event was described with a passive sentence; finally, (4) younger children provided more incorrect answers to our pictures than older children, regardless of whether the transitive action in the picture was described with an active or a passive sentence. In particular, the younger children provided significantly more incorrect answers to pictures with the verb "groeten" (to greet) compared to the older children.

Because our set contains many verb-noun combinations, the set is very useful to researchers who intend to investigate the acquisition of transitive structures in children in a longitudinal design, not only for studies with typically developing children, but also in studies with children who have a delay in their language development.
Keywords: Norming study, picture set, transitive structures, L1 syntax acquisition

Introduction

A crucial step in carrying out meaningful experiments is to choose appropriate stimuli. For instance, researchers in psychology and psycholinguistics often use pictures as stimuli in their experiments to investigate different aspects of language such as sentence production and comprehension. These pictures need to be explicit and clear enough such that they elicit specific syntactic structures from language users. A research domain in which pictures are often used to investigate the production and comprehension of syntax is the field of structural priming (Bock, 1986). Structural priming is the tendency of speakers to repeat structures that they have been recently exposed to; this is often investigated with picture descriptions or sentence completions (Mahowald et al., 2016). Most structural priming studies use pictures where transitive or ditransitive actions are displayed, and, often, these studies have used just two picture sets: the set by Bock (1986) and the one by Branigan et al. (2000). These sets have in common that they consist of black-and-white line drawings with combinations of animate and inanimate nouns performing or undergoing an action, and that they have been mostly used in studies with native speakers or highly proficient second language (L2) speakers (see e.g., Bernolet et al., 2009; Favier et al., 2019).

A difference between both sets is that the pictures[1] in Bock are rather outdated and may not necessarily appeal to, for instance, children, whereas the picture set of Branigan et al. seems more suitable to use in studies that investigate language production in special populations. However, both picture sets include verbs and nouns that are somewhat *infrequent*[2] and may be familiar to native language users or highly proficient L2 speakers, but not to other groups such as late L2 learners or children. Some infrequent verbs and nouns in Branigan et al. include, for instance, "to polish"[3], "to scold"[4], and "a monk"[5]. This important characteristic impedes researchers to confidently use the two picture sets when testing non-native or young language users.

Using frequent verbs (and nouns) in stimuli is especially important in studies involving children, since several studies have suggested that children are sensitive to lexical frequency information (Theakston et al., 2004); and that this frequency effect on the acquisition of lexical items is crucial in their development of syntactic knowledge (Matthews et al., 2005; see Ambridge et al., 2015 for an extensive review). If young children are exposed to an infrequent, and perhaps unfamiliar, lexical item and are asked to formulate a syntactic structure with that verb, they might have difficulty formulating the elicited syntactic structure since lexical representations play a role during the selection of syntactic structures (Ambridge et al., 2015; Pickering & Branigan, 1998). This highlights the importance of using frequent lexical items in stimuli that are aimed at investigating the production of syntax in young L1 users (but also in late L2 learners).

To keep the attention span of children to the task, many studies involving children use color drawings in their stimuli (see e.g., Huttenlocher et al., 2003; Goldwater et al., 2011). However, it is noticeable that the number of items in these studies is often limited. For instance, Huttenlocher et al. used 40 color drawings as their stimuli to investigate the priming of transitive and dative structures in children (two sets of 20 drawings for each structure). Moreover, Goldwater et al. used nine critical items (color drawings) in their structural priming experiment with 4- and 5-years old. Using a limited number of items usually requires more participants in the test sample to reach sufficient statistical power for observing priming effects (see Mahowald et al., 2016). It seems that, ideally, in studies with less experienced language users (i.e., children and late L2 speakers), there should be a balance between sufficient experimental items that use frequent verbs and enough participants.

Note that Muylle et al. (2020) designed a stimulus set of 423 animated movie clips depicting transitive and dative actions (prepositional object datives ["The woman gives a flower to the man"] vs. direct object datives ["The woman gives the man a flower"]) that can be used

---

[1] See OSF directory (https://tinyurl.com/DutchNormingStudy) for an example of pictures from both picture sets.

[2] In this paper, we denote frequent words based on Zipf-frequencies (Van Heuven et al., 2014; Keuleers et al., 2010). Zipf-frequencies range from 1 to 7, the values 1-3 indicate low frequent words (i.e., frequencies of 1 per million words and lower) and the values 4-7 indicate high frequent words (i.e., words with frequencies of 10 per million words or higher).

[3] "To polish" has a Zipf-frequency of 3.03

[4] "To scold" has a Zipf-frequency of 3.12

[5] "monk" has a Zipf-frequency of 3.86

to investigate transitive and ditransitive comprehension and production in groups such as adult L2 learners or children. Indeed, as the researchers propose, using clips that depict motion events may be easier for non-native or young language users since the action does not have to be inferred, and therefore, motion clips may be more ecologically valid than static pictures. However, some movie clips in Muylle et al. are rather violent (e.g., "A boxer punches a dancer"). Because of this, some of their clips may still not be very appropriate to use in studies with children.

In addition to animated movie clips, some studies have used live action objects, such as bunnies and ducks (Gertner et al., 2006) to elicit syntactic structures from young children in a language production task. Though these types of stimuli may strongly appeal to children, the development of such stimuli is effortful and can often only be used for the aforementioned group and not for other groups of language users (e.g., late L2 learners, people with aphasia...).

Here, we propose a set of colored pictures that was initially designed to test transitive comprehension and production in adult learners of Dutch. However, the purpose of the current study is to investigate whether our picture set is suitable to use in studies with young children too. For this reason, we conducted a norming study amongst young monolingual Dutch children.

Current study

In this study, we used 208 color drawings that were initially developed for longitudinal research on the production and comprehension of transitive sentences in adult learners of Dutch. Prior to constructing the color drawings for the late learners, we consulted their language learning materials to make sure that we used familiar verbs and nouns in our stimuli set[6]. Based on the learning materials, we chose transitive verbs (e.g., to call, to carry, to follow...), names of human professions (e.g., a cook, a baker, a teacher...), and names of vehicles (e.g., a car, a truck, a motorbike...) (see Appendix S1 for the verbs and nouns). We used a restricted number of verbs (12 verbs) and nouns (8 animate nouns and 8 inanimate nouns) for the color drawings since we wanted to avoid that the learners would be exposed to many different words. However, despite the small number of verbs and nouns that we used, our picture set was designed such that different verb-noun combinations are possible, which allows the same verb to be presented with different nouns while avoiding lexical overlap between items. This makes our picture set particularly suitable for longitudinal designs, since researchers can, for instance, test the acquisition of specific transitive verbs and both transitive structures (active and passive sentences) in adult and young language users, while avoiding a learning effect caused by repetition of the pictures.

A previous study showed that our picture set is indeed suitable to test the production of transitive structures in a group of late learners of Dutch. In a structural priming experi-

[6]To our knowledge, there exist no frequency data bases on the acquisition of Dutch verbs in late adult learners. For this reason, using the learners' learning materials was reliable for our study as it ensured us that we were using familiar words to these learners.

ment, Van Lieburg et al. (2022) observed that the learners produced 49.5% active sentence responses, 29.9% passive sentence responses and 20.7% "Other" responses to the transitive pictures. This suggests that the transitive events in our drawings largely elicit active as well as passive responses.

In the current study, however, we investigated whether the same picture set is suitable to use in studies involving young children. For this purpose, monolingual Dutch children (a group of 6-7 years old and a group of 11-12 years old) indicated whether our drawings can be interpreted equally well with an active and a passive sentence. We predict that these children will have difficulty interpreting our pictures when a transitive event is described with a passive sentence, since passives are complex structures and are learned relatively late (around the age of 3) compared to active sentences (Messenger & Fisher, 2018; Verrips, 1996). Passives are particularly complex to acquire by children due to their noncanonical word order (the agent of the sentence does not occur in the grammatical subject position compared to conventional word order rules [particularly in SVO languages]). Moreover, language production studies have suggested that children are more likely to first produce short passives (passive structures in which the by-phrase is not overtly realized: "The bear is hugged") than full passives (Fox & Grodzinsky, 1998). However, language comprehension studies have not found a consensus as to whether short passives are comprehended earlier than full passives (Hirsch & Wexler, 2006). Although young children can produce passive sentences spontaneously by the age of 3, they seem to have persistent trouble with understanding some type of passives. A common difficulty in young children's understanding of passives are reversed passives (Messenger et al., 2012; Armon-Lotem et al., 2016). Reversed passives are passives in which full grammatical knowledge is required for the child to interpret who does what to whom (Armon-Lotem et al., 2016). For instance, in the sentence "The woman was kissed by the man", the animate noun in the grammatical subject position and the one in the by-phrase can be exchanged since both nouns can do the kissing. Thus, here, the grammatical subject can be replaced with the agent in the by-phrase, while the sentence remains grammatically correct. If a child would be asked who does the kissing, they would most likely answer that the woman is the one doing the kissing (based on conventional word order rules). Nonreversible passives, however, are less challenging for young children to understand since children can rely on world knowledge to interpret who does what to whom (e.g., "The cheese was eaten by the mouse") (Armon-Lotem et al., 2016). Here, the child knows that only the mouse can do the act of 'eating'. Next to this, children have difficulty understanding passives with mental state verbs (e.g., *remembered, forget, know, believe* – "Joanna was forgotten by Olivia"). This type of passives is challenging for young children because the agent is experiencing the mental state (here, Olivia is doing the forgetting), while nothing "happens" to the patient. In contrast, passive sentences with actional verbs (e.g., *to push, to hug, to carry*...) are less challenging for children since the transitive action is carried out by the agent, while the patient undergoes the transitive action (Armon-Lotem et al., 2016). These features of the passive structure explain why the production and comprehension of (full) passives remains challenging throughout the first years of a child's education (Messenger et al., 2012).

Next to the difficulty in passive sentence comprehension, we also investigated to what extent the transitive verbs on our pictures would affect the children's comprehension of each transitive action, when an active or a passive was used to describe this action. We focused on verbs specifically (and not on the nouns in our pictures) because verbs are the most important building block in constructing syntactic structures (see Ambridge et al., 2015). Before conducting our experiment, we did not test whether the children knew the verbs that we used on our pictures, but we assumed that some of our pictures contain verbs that are abstract and may not have been acquired yet by the children. For example, one of our pictures depicts an event in which a car overtakes a truck ("De auto haalt de vrachtwagen in" – "The car overtakes the truck"), and the picture has been drawn from a bird's eye view (see Figure 1). Because adult learners have experience in traffic, understanding who passes whom is not particularly difficult. However, "inhalen" may be a challenging word for children because they do not have that same experience yet, especially because the action verb has been drawn from a bird's eye view[7]. What is more, young children might also have not acquired this verb yet ("inhalen" is acquired around the age of 8;61 years by Dutch children, see Brysbaert et al., 2014). In addition to this, "inhalen" has a Zipf-frequency of 2.72 (Keuleers et al., 2010); this means that it is considered a low frequent word[8], which may cause even more difficulty for children to interpret pictures with this verb correctly. Moreover, we expect that the combination of some abstract verbs (such as "inhalen") with a passive voice will further restrict the children's comprehension of the transitive event in our pictures.

Finally, we anticipate that, in general, the older children will be better at interpreting our pictures, either described with an active or a passive sentence, compared to the younger children. If this turns out to be the case, then our picture set will be suitable to test how children make progress in the comprehension of transitive structures.

Method

Participants

Two hundred participants (86 participants in the first grade (41 males and 45 females) and 114 participants in the fifth and sixth grade, 56 males and 58 females) took part in this study. Participants in the first grade were 6-7 years old and participants in the higher grades were 11 or 12 years old. Five classes in the first grade participated and five classes in the higher grades took part in this study (two classes in the fifth grade and three classes in the sixth grade). The participants were recruited at the same primary school in Antwerp, Belgium. All parents provided their consent for their child's participation in this study. The classes received a monetary reward for their participation. This study was approved by the ethical committee of the University of Antwerp (SHW_1877).

[7]Of course, if "inhalen" depicted a transitive event that is more familiar to children (a bike race or a running contest) or if pictures with this verb were drawn from a side view, the verb may be easier for children to comprehend. However, here, we specifically predict that the combination of the abstract verb "inhalen" together with the bird's eye view will probably require the children to make more inferences, which will probably lead to difficulty in comprehension.

[8]It is important to note that the verb Zipf-frequencies in Keuleers et al. (2010) are based on Dutch film subtitles.

Stimuli & Design

Stimuli

We used 208[9] pictures for the present study. The 208 pictures were pairs of target and competitor pictures (thus, there were 104 pairs). We used 12 different verbs to depict transitive events in the set of 208 pictures. The target and competitor pictures always used the same verb. The verbs were more or less equally divided across the 208 pictures: four verbs (*aanrijden* [to hit], *bellen* [to call], *dragen* [to carry], and *volgen* [to follow]) each occurred 20 times in the set and eight other verbs (see Appendix S1 for all the verbs; their Zipf-frequency and the age of acquisition of these verbs) occurred 16 times in the pictures. Importantly, because we used the verbs based on the L2 learning materials of late Dutch learners (see above), we did not control for how many verbs were abstract or concrete (i.e., we did not equally divide the verbs into an abstract and concrete category). We believe that our materials consist of 3 abstract verbs (*inhalen* [to overtake], *vervangen* [to replace]; *blokkeren* [to block]), while the remaining 9 verbs may be considered concrete verbs (see Appendix S2 for examples of pictures with the abstract verbs "vervangen" and "blokkeren"; see Figure 1 for an example with the verb "inhalen"). Next to this, the agents and the patients of the transitive actions could be one of eight animate entities referring to a profession (e.g., a baker, a construction worker, a cook) or one of eight inanimate entities referring to a vehicle (e.g., an ambulance, a firetruck, a car [see Appendix S1 for all the nouns]). Thus, our sentences contained only animate subject – animate object (e.g., "De bakker helpt de zangeres" – "The baker helps the singer") or inanimate subject – inanimate object ("De vrachtwagen volgt de auto" – "The truck follows the car") combinations. It is important to note that all sentences in our picture set were reversible (i.e., the agent and patient can be exchanged while our sentences remain grammatically correct) and that all our verbs were actional verbs (our materials did not have any mental state verbs).

Design

The target-competitor pairs were divided in five different lists consisting of 42 or 41 items. After doing this, we randomly assigned the five classes in the lower and higher grade to one of the five lists (these lists were later converted into separate booklets, see below). The first three lists had 42 target-competitor pairs and the last two lists consisted of 41 pairs. In every list, we manipulated the position of the target pictures within each target-competitor pair. This means that, within a list, half of the target pictures (21 or 20 pictures) appeared on the left and the accompanying competitor pictures appeared on the right (see Figure 1). Similarly, the other half of the target pictures (also 21 or 20 pictures) were on the right and the accompanying competitors were on the left. Next to this, we aimed to equally present an active and passive interpretation of a target picture within each target-competitor pair.

[9]There were 194 unique pictures, and 14 pictures were used twice, but elicited two different states of affairs (e.g., the picture for an ambulance replacing a bus elicited the sentence "the ambulance replaces the bus" in one trial, while it elicited the sentence "the bus is being replaced by the ambulance" in another trial). The pictures that occurred twice in the set are marked in the OSF directory.

For this reason, within a list, half of the sentences (i.e., the sentences that participants would hear during the experiment) were actives and the other half were passives. For example, if the children would be presented with an active sentence “The car overtakes the truck”, they would see two pictures (see Figure 1). In the correct target picture, the agent is the car, and the truck is the patient. In the competitor picture, the agent is the truck, and the patient is the car. The competitor picture could either be described with an active sentence “The truck overtakes the car” or with a passive sentence “The car is being overtaken by the truck”. We included the competitor picture to test whether the children would interpret the first noun phrase as an agent (active interpretation) or a patient (passive interpretation).

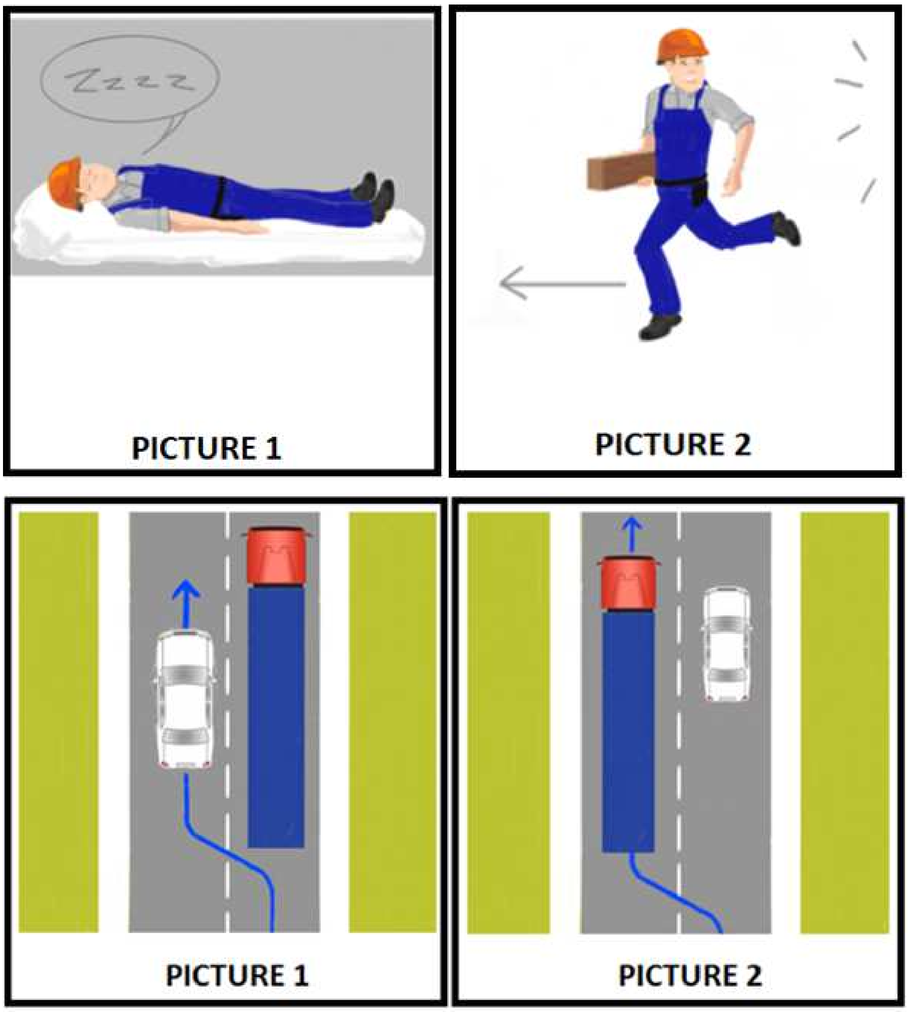


Figure 1: An example of a target-competitor pair in the practice trial for the sentence “De bouwvakker slaapt” – The construction man sleeps (1) and in an experimental trial (2) for the sentence “De auto haalt de vrachtwagen in” – The car overtakes the truck. For both trials, the correct picture is on the left and the competitor picture is on the right.

Because each target-competitor item appeared once within a list, we could only present the active or passive interpretation of that target picture. Since our goal was to test the children's comprehension of a target picture in both transitive sentence structures, we manipulated the condition of each target picture across lists. That is, each target picture was presented with an active sentence in one list and with a passive sentence in another list. In this way, we collected data on whether children interpreted the target picture correctly when it was described with an active sentence and its passive counterpart. Furthermore, the target-competitor items were constructed such that each transitive verb occurred at least once in each list; thus, every participant was exposed to all twelve verbs.

Our list construction deviates from conventional experimental lists in psycholinguistics due to several reasons. (1) Because we tested young children, we could not expose them to all 104 target-competitor pair items due to their attention span and thus, we reduced the number of items that children were exposed to within a list (42 or 41 items). (2) Our aim was to test whether a target picture could be interpreted equally well with an active and passive sentence, and because of this, across lists, there were two conditions, but we had five experimental lists (the first 3 lists had 42 items, and the last two list had 41 items). For this reason, we selected a different set of items for each list.

Procedure

Because of the COVID-19 pandemic, the experiment could not be administered by the researchers themselves. For this reason, the children were tested in class and the experiment was administered by their teachers, whom we provided with detailed instructions. Each class was assigned a separate list (there were five lists in total). We presented the five lists with the target-competitor items in five separate Microsoft PowerPoint presentations. The teachers were also provided with a Word Document containing the pictures of the (in)animate nouns that would appear in the experiment. The teachers first showed this document on a digital white board to the children to familiarize them with the pictures of the nouns and their accompanying names. After this, the teachers could start with the experiment, which was presented in a PowerPoint Presentation on a digital white board. Each list (i.e., each presentation) started with the same practice trial consisting of an audio file of the intransitive sentence "De bouwvakker slaapt" – "The builder sleeps". We used an intransitive sentence in the practice trial so that the goal of the experiment (how children interpret transitive events when a passive or an active sentence is used) would not be revealed. After participants had listened to the audio file, the teacher showed the target and competitor pictures describing the intransitive event on a separate page in the PowerPoint presentation (see Figure 1). The two pictures were framed with a black line. The phrase "picture 1" appeared at the bottom of the left picture, and "picture 2" appeared at the bottom of the right picture (Calibri bold, 18-point font, see Figure 1). At this point, the teacher asked the class which picture depicted the sentence "The builder sleeps" and the children had to say the answer out loud. After the practice trial, teachers could start with the experimental trials. For each trial, the children listened to an audio file of a sentence describing either an active or passive transitive event. Subsequently, a picture of the target and competitor item, that depicted the transitive event, was shown on the PowerPoint-presentation. Crucially, the same target-competitor items that were shown in the PowerPoint presentation were also presented to each participant in their answer booklet. It is important to make clear that, unlike the practice trial that did not appear in participants' answer booklet, the experimental trials all appeared in the children's answer booklet. In each child's answer booklet, every target and competitor item appeared on one page next to each other (see Figure 1 for an example of the verb 'inhalen' – to overtake). The presentation of the target-competitor items in the PowerPoint presentation and in participants' answer booklet was done to enable group testing. Participants had to indicate which picture (either picture 1 or 2) matched the sentence they had heard; they marked this in their booklets.

We told teachers that each audio file could be listened to twice. Because we wanted to provide teachers with some flexibility in how they played the audio files to the children, teachers could either choose to let the children listen to an audio file the first time, after which they marked the picture in their booklet, and listened to the same audio file again, or teachers could let the children listen to each audio file twice before they provided an answer in their booklet. In this way, teachers could freely choose which method worked best for their class. We emphasized that once teachers had chosen a method, they had to stick to it throughout the experiment. We also provided teachers with an Excel sheet in which the audio sentences were written down. Teachers could consult the sheet in case they wanted to know which sentence the children would hear next. We instructed teachers to allow the children to ask questions during the experiment or to further explain what was expected from them. Also, we told teachers that if any technical difficulties would occur during the experiment (i.e., if an audio file could not be played), they were allowed to read the sentence out loud to the children; they could use the Excel sheet for this. Sessions took approximately 30 to 45 minutes.

Coding

The answers to the pictures were coded as correct or incorrect. If the target picture was marked, the trial was coded as correct and if the competitor was marked instead, the trial was coded as incorrect. Moreover, if it was clear that the pupils first gave one answer and then changed it (e.g., by erasing it), the first given answer was coded. We did this because, in this study, we were specifically focused on the children's initial comprehension of the sentence they had heard. We believe that including the self-corrected trials in the analyses would not provide us with the most accurate comprehension of a given picture. On trials where children did not provide an answer, the trial was marked as a missing value (i.e., NA). We did not include the missing values in our analyses reported below, since our dependent variable was a binary response, and the number of missing values was very low (see below).

Analysis

We fitted the answers to the pictures (correct answer vs. incorrect answer, correct answer as the reference level) to generalized linear mixed effect models as implemented in the *lme4* package (Bates et al., 2015) in R (R Core Team, 2020). Our full model consisted of a three-way interaction between *Condition* (active, passive, active as reference level) * *Verb* (aanrijden [to hit] as reference level) * *Grade* (high grade vs. low grade, high grade as reference level); we used a Bobyqa optimizer to increase convergeability (Powell, 2009). We accounted for the non-independence between observations from the same participant and from the same item, by entering the random intercepts for *participant* and *item*. Conform the maximal random effects structure as proposed by Barr et al., (2013), we added *Condition* and *Verb* as random slopes over *participant* and *Condition* as a random slope over *item*.

Due to convergence issues and singular fit warnings, the maximal model was simplified in a stepwise way. We first simplified the random effects structure by testing if the random slope terms could be omitted without decreasing the fit of the model. Second, the fixed effects part was simplified by testing which interactions were not significant; for this, we used the *drop1* function from the basic stats package in *R* (version 4.1.2) Based on the output that we received from the drop1 function, we decided which interactions could be dropped from the model (*Condition** *Verb** *Grade*, and *Condition** *Grade* were not significant). Lastly, the conditional and marginal R^2 values of the final model, which are measures of effect sizes, were calculated using the *rsquared* function from the piecewiseSEM package (version 2.1.0., Lefcheck, 2016).[10]

Results

In total, we collected 8,319 answers to the active and passive descriptions of our pictures. There were 7,700 correct answers (92.60%), 578 incorrect answers (6.95%) and 41 missing values (0.493%).

For the active condition, the *higher* grade provided 2321 correct answers (97.8%), 39 incorrect answers (1.64%) and there were 11 missing values (0.46%). For the passive condition, the higher grade provided 2272 correct answers (95.8%), 93 incorrect answers (3.92%) and they did not give an answer on 6 trials (0.25%).

For active sentences, the *lower* grade gave 1649 correct answers (92.2%), 136 incorrect answers (7.61%) and there were 3 missing trials (0.17%). For passive sentences, the lower grade provided 1458 correct answers (81.5%), 310 incorrect answers (17.3%) and there was no answer for 21 trials (1.17%). Figure 2 shows the number of correct and incorrect answers per sentence condition per grade. However, note that the percentages are slightly different from the percentages listed above since we did not include the missing values in the plot and in further analyses.

Our final model consisted of a significant two-way interaction between *Condition** *Verb* ($p < .001$) and a two-way interaction between *Verb** *Grade* ($p < .001$) and *participant* and *item*

[10]We performed exploratory analyses to investigate whether there was an interaction between Zipf-frequency * Verb and an interaction between AoA (age of acquisition) * Verb, but we found no interaction between these variables. For this reason, we decided to exclude the Zipf-frequencies and AoA values from our main analysis (see Appendix S1 for these values per verb). Throughout the rest of the paper, further comments on verb frequencies and AoA are thus rather descriptive than drawn from inferential statistics.

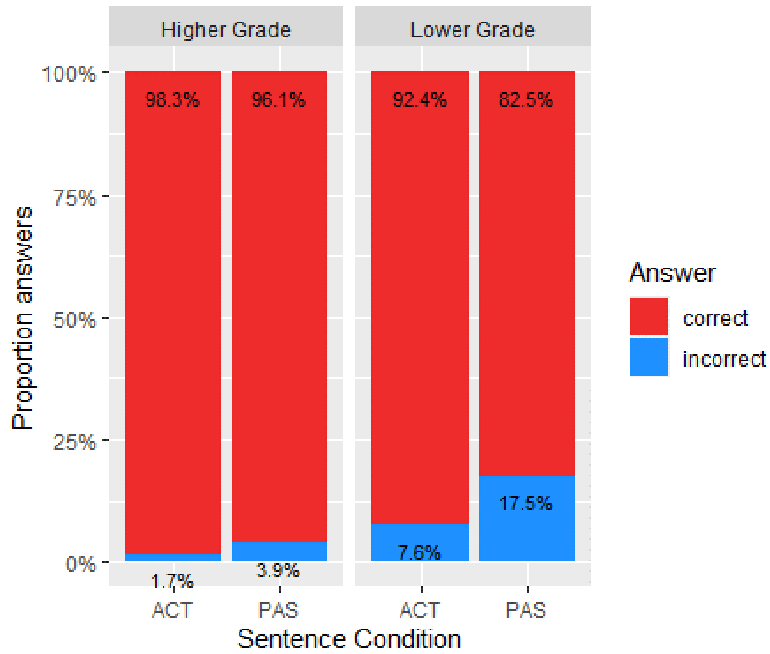


Figure 2: Proportion of correct and incorrect answers per sentence condition and per grade.

as random intercepts. The fixed effects of the final model explained 14.84% of the variance (marginal R^2; Nakagawa & Schielzeth, 2013) and conditional on the random effects, they explained 21.10% of the variance (conditional R^2; Nakagawa & Schielzeth, 2013). Table 1 only shows the significant results in the output for the final model.

The negative estimate for the intercept indicates that children provided significantly more correct answers (β = -5.58, $p < .001$) to active sentences than to passive sentences. There was thus a significant increase in the number of incorrect answers for passive sentences compared to active sentences, which is noticeable in the strong effect for the predictor *Condition* ($\beta = 1.55$, $p < .001$).

For the verbs, we found a significant effect for "inhalen" (to overtake): the positive estimate suggests that this verb significantly rendered more incorrect answers to the target pictures in comparison to the reference verb "aanrijden" (to hit) in active sentences (β = 2.37, $p < .001$). We found the same effect for the verb "vervangen" (to replace) ($\beta = 2.69$, $p <$ .001).

The positive estimate for the predictor *Grade* (with high grade as reference level) suggests that children in the lower grade significantly provided more incorrect answers in the reference level than children in the higher grade (β 1.80, $p < .001$) (see also Figure 3 for the percentages).

Though there was a significant interaction between *Condition* * *Verb* (p < .001) (obtained by the *drop1* function), this was not the same for all levels of the predictor *Verb*. That is, we found a significant interaction between *Condition* and the verb "blokkeren" (to block");

Table 1: Final model output with the predictors *Condition* (active condition as reference level), *Verb* (aanrijden as reference level) and *Grade* (high grade as reference level). We only report the significant results in this table (see Appendix S3 for the full model output).

Summary of the fixed effects in the multilevel logit model (N = 8278; **log-likelihood** = -1594.9) **Fixed effects**	β-coefficient	SE	Z-value	p-value
(Intercept)	-5.58	0.51	-10.89	<.001***
Condition (passive)	1.55	0.40	3.86	<.001 ***
Verb (inhalen)	2.37	0.58	4.06	<.001 ***
Verb (vervangen)	2.69	0.58	4.62	<.001 ***
Grade (low)	1.80	0.41	4.37	<.001 ***
Condition (passive) * Verb (blokkeren)	-1.89	0.56	-3.37	<.001 ***
Condition (passive) * Verb (inhalen)	-1.12	0.46	-2.41	<.05 *
Condition (passive) * Verb (roepen)	2.11	1.06	1.98	<.05 *
Condition (passive) * Verb (volgen)	2.30	0.81	2.82	<.01 **
Verb (groeten) * Grade (low)	1.52	0.70	2.18	<.05 *

the negative estimate suggests that, in contrast with the active condition, the number of incorrect answers for this verb in the passive condition was lower than for the reference verb “aanrijden” (to hit) (β = -1.89, $p < .001$) (see Figure 3). Moreover, the negative estimate of the interaction between *Condition* and the verb “inhalen” (to overtake) suggests that the effect of condition was smaller for this verb than for the reference verb (β = -1.12, $p < .05$). We also found a significant interaction between *Condition* and the verb “roepen” (to call) (β = 2.11, $p < .05$). As the estimate is positive, the effect of Condition is larger for this verb than for the reference verb “aanrijden”. We observed the same effect for the significant interaction between *Condition* and the verb ‘volgen’ (to follow) (β = 2.30, $p < .01$): here, the increase in the number of incorrect responses in the passive condition was also larger in comparison with the reference verb (see Figure 3).

There was also a significant interaction between the verb “groeten” (to greet) and *Grade*; the positive estimate indicates that children in the lower grade provided more incorrect answers for this verb compared to the reference level than children in the higher grade (β = 1.52, $p < .05$).

Note that we did not find a significant interaction between *Condition***Grade*, though, descriptively, it seems that children in the lower grade provided more incorrect answers in passive sentences than in active sentences compared to children in the higher grade (see Figure 3). Due to the variance in the many words that we used, the interaction between our condition and the two grades might have not been statistically significant.

Because the effect for each verb on the outcome variable was compared to only one verb

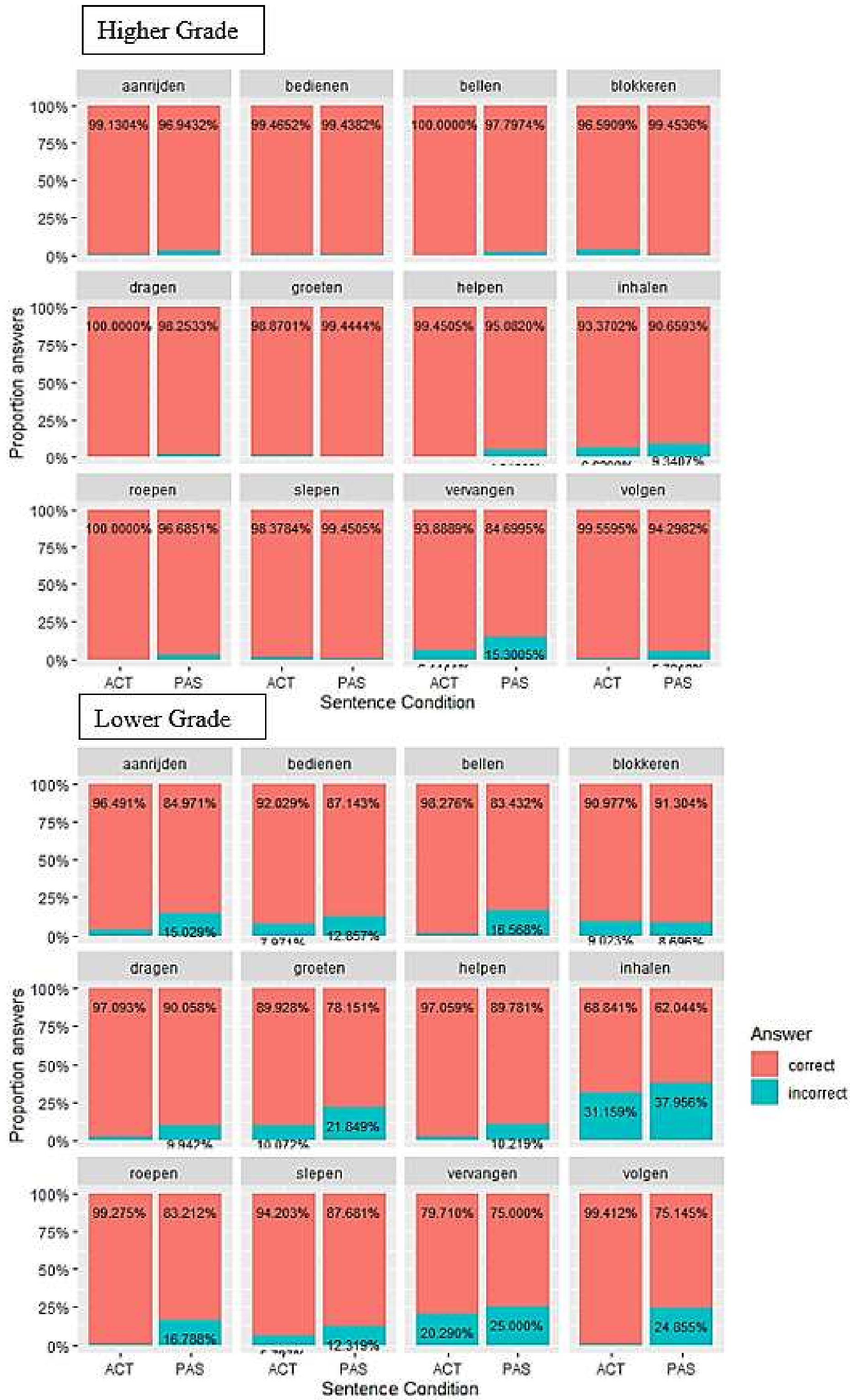


Figure 3: The proportion of correct and incorrect answers to each verb per sentence condition per grade. The upper bar plot shows the higher grade and the lower bar plot the lower grade. The pink bars indicate the correct answers and the blue bars the incorrect answers.

(the reference level “aanrijden” [to hit]), we performed pairwise comparisons using the *emmeans* function (from the package ‘emmeans’, Lenth 2019) to investigate whether the verbs significantly differed from each other (see Table 2). From Table 2, it seems that the verbs “inhalen” (to overtake) and “vervangen” (to replace) were indeed more difficult to comprehend for the children compared to the other ten verbs (see also Figure 3).

Table 2: Pairwise comparisons of the verbs. In the first column, the first verb significantly rendered more correct answers than the second verb. In the second column, the second verb produced more correct answers than the first verb.

More **correct** answers for the **first** verb than for the **second** verb		More **correct** answers for the **second** verb than for the **first** verb	
aanrijden – inhalen:	$Z = -5.47$, $p < .001$	inhalen – slepen:	$Z = 5.22$, $p < .001$
aanrijden – vervangen:	$Z = -5.19$, $p < .001$	inhalen – roepen:	$Z = 4.92$, $p < .001$
bedienen – inhalen:	$Z = -5.07$, $p < .001$	vervangen – volgen:	$Z = 4.92$, $p < .001$
bedienen – vervangen:	$Z = -4.85$, $p < .001$		
bellen – inhalen:	$Z = -5.87$, $p < .001$		
bellen – vervangen:	$Z = -5.65$, $p < .001$		
blokkeren – inhalen:	$Z = -4.72$, $p < .001$		
blokkeren – vervangen:	$Z = -4.46$, $p < .001$		
dragen – inhalen:	$Z = -6.27$, $p < .001$		
dragen – vervangen:	$Z = -6.02$, $p < .001$		
groeten – inhalen:	$Z = -4.41$, $p < .001$		
groeten – vervangen:	$Z = -4.18$, $p < .01$		
helpen – inhalen:	$Z = -5.29$, $p < .001$		
helpen – vervangen:	$Z = -5.04$, $p < .001$		
roepen – vervangen:	$Z = -4.74$, $p < .001$		
slepen – vervangen:	$Z = -4.96$, $p < .001$		

Discussion

In this study, a group of 200 Dutch monolingual Dutch children indicated whether our 208 color drawings depicting transitive action events could be interpreted equally well with an active and a passive sentence. Generally, we found that the children provided 93.02% correct answers to the pictures when an active or a passive sentence was used to describe a transitive event in a picture. In total, there were only 6.98% incorrect answers to the pictures. These results imply that our pictures can be interpreted with both transitive structures and are suitable for language studies with (Dutch) children.

Though the children performed well on the task, we specifically investigated what factors contributed to incorrect answers. This is important to know as these factors may provide valuable information for other researchers who want to use our picture set in their study. There were several factors that may have contributed to incorrect answers: (1) the use of an active or a passive sentence to describe a picture; (2) the verb that was used to describe the transitive action in a picture; (3) and the age of the children that we tested (we tested children in two different grades, a lower [6-7 years] and a higher grade [11-12 years]).

The strong main effect for our predictor *Condition* indicates that the children had more difficulty interpreting a picture when a passive sentence was used to describe a transitive event than when an active sentence was used. From early on, children receive little input to learn the passive structure. For instance, Gordon and Chafetz (1990) only found four full English passives in a corpus of 87,000 child-directed utterances. Moreover, it takes a while before school-age children (6 years and older) have acquired an adult-like interpretation of complex structures, such as passives (Montgomery et al., 2017). More importantly, all of our sentences were reversible passives. That is, the agent and patient of the passive structure can be exchanged while the sentence remains correct (e.g., “De bakker wordt gegroet door de leraar” - “The baker is being greeted by the teacher”. As mentioned in the introduction, reversible passives are challenging for children because they have difficulty interpreting who does what to whom (see for a review Armon-Lotem et al., 2016). Non-reversible passives are easier to comprehend because children can rely on world knowledge to infer the meaning of the passive structure (e.g., The gras is being eaten by the cow). Researchers who intend to use our materials to investigate transitive comprehension in (Dutch) children should thus be aware of the fact that our pictures only consist of reversible passives. If our materials included non-reversible passives, it is possible that we might have observed a less strong effect for *Condition*.

We also tested to what extent the verb, describing the transitive event on our pictures, affected the number of incorrect interpretations of the transitive action. We found that the verbs “inhalen” (to overtake) and “vervangen” (to replace) produced significantly more incorrect answers to the pictures compared to our reference verb “aanrijden” (to hit) in active sentences. There could be several explanations for why these two verbs were significantly more difficult for children to comprehend than the other ten verbs that we used in our stimuli. We believe that the combination of these verbs together with how they were depicted on our pictures might have forced the children to make more inferences than on pictures that depicted concrete verbs (e.g., the verb “dragen” [to carry]). For instance, pictures with the verb “inhalen” were drawn from a bird’s eye view perspective (see Figure 1 for an example). The bird’s eye view may have hindered children’s comprehension of “inhalen” because it might have been difficult to parse what is overtaking what. Even though we used arrows to indicate what object was performing the act of overtaking another object, it is likely that using arrows to indicate a motion might have been too abstract for the children. In this case, they do not only have to infer the meaning of the verb, but also the meaning of a “static” arrow that should depict a motion. The same argumentation can be made for the verb “vervangen” (see Appendix S2). Not only is “vervangen” a rather abstract verb, its depiction on our pictures may have also contributed to difficulties in processing the transitive event. More specifically, we depicted the transitive action of “replacing” with a red cross. For instance, for the sentence, “De ambulance vervangt de bus” [“The ambulance replaces the bus"], we put a red cross through "the bus"; this was supposed to indicate that this object was replaced by the other object that did not have this cross (see Appendix S2). It is very likely that the combination of the picture and the verb may thus have caused processing difficulties for the children. Moreover, it could be possible that the children in our study might not have been familiar with both verbs, as these verbs are learned relatively late ("inhalen" is learned at 8;61 years and "vervangen" at 9;06 years, see Brysbaert et al., 2014). In our exploratory analyses, where we tested whether there was an interaction between *Verb* and AoA (age of acquisition), we did not find a significant interaction between these two variables. However, we still believe that some children (especially the younger ones) might not have acquired these two verbs yet at the time of testing. For this reason, this might have increased effortful processing of the pictures that contained these two verbs.

We also anticipated that comprehending some verbs could be even more difficult when they were used in passive sentences. Two verbs produced more incorrect answers to the pictures when a passive sentence was used compared to the reference verb "aanrijden" (to hit) in active sentences, namely, "roepen" (to shout) and "volgen" (to follow). Although "roepen" and "volgen" are learned somewhat early by Dutch children, ("roepen" is learned at the age of 4;96 years and "volgen" is learned at 5;84 years, Brysbaert et al., 2014), there could be other explanations for why these verbs produced significantly more incorrect answers to the pictures when they were used in a passive sentence. One could argue that the verb "roepen" (to shout) is a so-called *action verb*, which implies that this verb, ideally, appears in a sentence where the agent is in the subject position and the patient is in the direct object position of a sentence (Lempert, 1989). In other words, it may be that this verb is tightly linked to an argument structure that is compatible with how an active sentence is formed. Although "roepen" can equally occur in an active and a passive sentence, the verb's syntactic preference (i.e., verb bias, see Peter et al., 2015) may affect which syntactic structure children (and other speakers) expect the verb to occur in. What is more, it is likely that children are more frequently exposed to this verb in active sentences than in passive sentences. Thus, when "roepen" occurred in the unexpected passive sentence, the children might have had difficulty to parse who called whom in the picture (i.e., it is likely that the first noun was interpreted as an agent instead of a patient, leading to an incorrect answer for the passive interpretation of a picture). For "volgen" (to follow), the situation is a bit different. We believe that the pictures containing this verb might have caused some confusion for the children. In the pictures with the verb "volgen", we used an arrow to indicate what followed what. For instance, for the sentence "the car follows the truck", we used an arrow to display the direction of the transitive event (an arrow was placed between the agent and the patient, see Appendix S3). The arrow might have caused participants to automatically interpret the first noun as the agent ("the car") followed by the patient ("the truck"). In the case that the picture was described with a passive instead, children had to disregard the direction of the arrow, since the patient was mentioned first. For this reason, we assume that, since the arrow might have prompted an active interpretation of the transitive event, children might have had difficulty to comprehend what followed what when a passive sentence was used, leading to more incorrect answers for passives as opposed to actives.

We found the opposite effect for the verbs "blokkeren" (to block) and "inhalen" (to overtake): it seems that these verbs produced significantly more correct answers in passive sentences compared to our reference verb “aanrijden” (to hit) in active sentences. The pictures in which these verbs were depicted always consisted of an inanimate agent and an inanimate patient (e.g., “De vrachtwagen haalt de auto in” – The truck overtakes the car; “De vrachtwagen blokkeert de auto” – The truck blocks the car). Several studies (see Bock, 1986; Gàmez & Vasilyeva, 2015) have shown that people have the tendency to produce passive sentences over active sentences when they encounter an inanimate actor and an animate undergoer. In our case, sentences with both transitive verbs (“blokkeren” and “inhalen”) had an inanimate agent and patient. However, it could be the case that when children encountered the first inanimate noun, they might have activated a passive sentence over an active sentence, since the grammatical subject of the sentence was inanimate. Possibly, this could have facilitated the processing of pictures that contained these verbs.

Lastly, as expected, we found that the children in the lower grade provided more incorrect answers to the pictures (either described with active or passive sentences) than the children in the higher grade. This means that the older children were better at matching our pictures to the correct transitive structure (i.e., older children had a better comprehension of transitive structures than younger children). Additionally, we found a significant interaction between the verb “groeten” (to greet) and *Grade*: children in the lower grade significantly made more errors in interpreting pictures with this verb in active sentences compared to children in the higher grade. An explanation for this could be that “groeten” is a rather formal word that young children would not necessarily use themselves. More importantly, “groeten” is acquired at the age of 6;71 years (see Brysbaert et al., 2014), which implies that the children in the low grade might not have been familiar with this verb yet (or they might be familiar with this verb only in specific contexts), while the older children might have already known this verb. Another explanation could be that the transitive action of this verb was probably not clear on its pictures. The pictures with the verb “groeten” have two animate objects where the agent is holding one hand up (indicating the action of “to greet”) while the patient is portrayed statically on the picture. It could be that the action event of the agent greeting the patient was not depicted explicitly enough, since the only indication for it is the agent’s hand being held up. Children in the lower grade might have overlooked this important feature while trying to comprehend pictures with the verb “groeten”, leading to incorrect answers.

Limitations and Future Directions

Though we chose to construct our picture set for Dutch language users, since, to our knowledge, no such set exists yet, we admit that this also comes with some disadvantages. As Dutch is only spoken in a few countries (the Netherlands, Belgium, and former Dutch colonies, such as Surinam, Curaçao, the country of Sint Maarten and Aruba), the use of this picture set may be limited. Of course, the transitive verbs can be translated to English (and other languages), which allows these pictures to be used in studies involving speakers of English. However, ideally, before using it in studies with, for instance, English speakers, one would first have to investigate whether the English variant of our picture set produces approximately the same number of correct interpretations of the transitive events in our pictures (either described with an active or a passive sentence).

We are aware that the current study addresses verb frequency and the age of acquisition of the verbs descriptively. Because our materials were initially developed for late L2 learners of Dutch, we used verbs that were familiar to them. Prior to conducting this study, we did not test whether the children also knew these verbs and whether they had already acquired them or not (as we were mainly interested in whether children were able to interpret our pictures equally well with an active and a passive structure). For this reason, if researchers intend to use our materials to test, for instance, the online processing of transitive structures, it should be noted that lexical frequency may affect sentence processing (Huizeling et al., 2022). We emphasize, however, that the children only provided 7% incorrect answers, which means that we should not overlook the fact that our materials did largely elicit correct interpretations of both active and passive descriptions of transitive events (93%).

Another possible limitation is that we used our picture set to investigate the comprehension of active and passive events in a group of children who were already relatively far into their L1 syntax acquisition. As we know that children start understanding passives around the age of 3, we do not know whether this picture set is also suitable for the youngest group of language users. Because of the findings in the current study (it turns out that our picture set contains a few transitive verbs that are learned quite late by Dutch children), we believe that our picture set might probably be too complex for children younger than the age of 6. For this reason, we advise researchers to use our picture set for children in the same age group as the one we tested (6 – 12 years old).

Throughout this article, we used studies in structural priming involving children as a possible research domain for which our picture set can be used. However, we would like to stress that our picture set may be suitable for other research domains too, particularly in domains where sentences are elicited. Since our picture set was initially designed for longitudinal designs, our set can be useful to test the acquisition of transitive structures in children, especially the use of passives. Research has shown that passive sentences are quite difficult for children to such an extent that, even when they have already learned the passive structure, their passive sentence production remains effortful (Messenger et al., 2012). As our norming study showed that the younger children made more errors in *comprehending* pictures that were described with a passive sentence compared to the older children, future studies could investigate whether younger children also make more errors while *producing* passive sentences, and at what age the number of errors decreases. Apart from using our picture set to investigate the acquisition of transitives in typically developing children, our stimuli can also be used in studies that involve children with a delayed language development (e.g., children on the autism spectrum) as acquiring passives has been found to be even more effortful for them compared to children without such a delay (Ambridge et al., 2021). Moreover, because our picture set uses a limited vocabulary (verbs and nouns), we believe that it may not be too taxing on the language abilities of, for instance, people with a language impairment. We also think that, for the same reason, our picture set can be used in studies with healthy elderly people.

Lastly, we tested the comprehension of transitive structures, but our pictures could also be used in other modalities such as language production (spoken as well as written, see Van

Lieburg et al. [2022], who already used our picture set to investigate L2 syntax production). An interesting question could be whether language users spontaneously describe a transitive event with the more complex passive sentence structure or whether they will almost always produce active sentences because they are easier. This question is especially interesting for written language production, as research has shown that people use passives more spontaneously during writing than in spoken language (Hinkel, 2004).

Conclusion

This study set out to test whether our picture set, consisting of transitive events with different verbs, could be interpreted equally well with active and passive sentences amongst Dutch children who differed in age. Even though the children provided 93.02% correct answers, we found that they provided significantly more incorrect answers to pictures that used a passive sentence to describe a transitive event than to pictures that used an active sentence. Moreover, we found that some verbs ("inhalen" and "vervangen") significantly produced more incorrect answers regardless of whether they were presented in an active or a passive sentence compared to the other verbs that occurred in our pictures. Moreover, we observed that the children had difficulty in comprehending some verbs when they were presented in a passive sentence, while other verbs rendered more correct answers when the transitive event was described with a passive sentence. We also found that, in general, younger children had more difficulty in interpreting our pictures, either described with an active or a passive sentence. Only one verb seemed to be particularly difficult for the younger children, namely, the verb "groeten" (to greet). We believe that the relatively late age of acquisition of this verb (6;71 years) and the unclear depiction of the transitive action "groeten" on pictures could have hindered the young children's comprehension of this verb.

Altogether, our study shows that the colored pictures are mostly clear in terms of how the transitive events are displayed and how well they can be interpreted with either an active or a passive sentence by young children.

Transparency and Openness

The color drawings, all data, analysis code, and experimental lists are available at https://tinyurl.com/DutchNormingStudy. This study's design and the analyses were not pre-registered.

Acknowledgments

This research was funded by a BOF DOCPRO4 grant from the University of Antwerp (nr. 37003).
OSF repository: https://tinyurl.com/DutchNormingStudy.
We have no known conflict of interest to disclose.

Literature

Ambridge, B., Bidgood, A., & Thomas, K. (2021). Disentangling syntactic, semantic and pragmatic impairments in ASD: Elicited production of passives. *Journal of child language, 48*(1), 184-201. https://doi.org/10.1017/S0305000920000215

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of child language, 42*(2), 239 273. https://doi.org/10.1017/S030500091400049X

Armon-Lotem, S., Haman, E., Jensen de López, K., Smoczynska, M., Yatsushiro, K., Szczerbinski, M., ... & van Der Lely, H. (2016). A large-scale cross-linguistic investigation of the acquisition of passive. *Language acquisition, 23*(1), 27-56.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1- 48. https://doi.org/10.18637/jss.v067.i01

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2009). Persistence of emphasis in language production: A cross-linguistic approach. *Cognition, 112*(2), 300- 317. https://doi.org/10.1016/j.cognition.2009.05.013

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13-B25. https://doi.org/10.1016/S0010-0277(99)000815

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica, 150*, 80-84. https://doi.org/10.1016/j.actpsy.2014.04.010

Czypionka, A., Spalek, K., Wartenburger, I., & Krifka, M. (2017). On the interplay of object animacy and verb type during sentence comprehension in German: ERP evidence from the processing of transitive dative and accusative constructions. *Linguistics, 55*(6), 1383-1433. https://doi.org/10.1515/ling-2017-0031

Favier, S., Wright, A., Meyer, A., & Huettig, F. (2019). Proficiency modulates between- but not within-language structural priming. *Journal of Cultural Cognitive Science, 3*(1), 105–124. https://doi.org/10.1007/s41809-019-00029-1

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language, 33*(6), 715-736. https://doi.org/10.1006/jmla.1994.1034

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied Longitudinal Analysis.* John Wiley & Sons.

Fox, D., & Grodzinsky, Y. (1998). Children's passive: A view from the by-phrase. *Linguistic inquiry, 29*(2), 311-332.

Gámez, P. B., & Vasilyeva, M. (2015). Exploring interactions between semantic and syntactic processes: The role of animacy in syntactic priming. *Journal of Experimental Child Psychology, 138*, 15-30. https://doi.org/10.1016/j.jecp.2015.04.009

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological science, 17*(8), 684 691. https://doi.org/10.1111/j.1467-9280.2006.01767.x

Goldwater, M. B., Tomlinson, M. T., Echols, C. H., & Love, B. C. (2011). Structural priming as structure-mapping: children use analogies from previous utterances to guide sentence production. *Cognitive science, 35*(1), 156-170. https://doi.org/10.1111/j.1551-6709.2010.01150.x

Gordon, P., & Chafetz, J. (1990). Verb-based versus class-based accounts of actionality effects in children's comprehension of passives. *Cognition, 36*(3), 227-254. https://doi.org/10.1016/0010-0277(90)90058-R

Hinkel, E. (2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language teaching research, 8*(1), 5-29. https://doi.org/10.1191/1362168804lr132oa

Hirsch, C., & Wexler, K. (2006). Children's passives and their resulting interpretation. In *The proceedings of the inaugural conference on generative approaches to language acquisition–North America, University of Connecticut Occasional Papers in Linguistics* (Vol. 4, pp. 125-136). Storrs, CT: University of Connecticut.

Huizeling, E., Arana, S., Hagoort, P., & Schoffelen, J. M. (2022). Lexical frequency and sentence context influence the brain's response to single words. *Neurobiology of Language, 3*(1), 149-179.

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of memory and language, 50*(2), 182-195. https://doi.org/10.1016/j.jml.2003.09.003

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior research methods, 42*(3), 643-650.

Lefcheck, J.S. (2016). "piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics." *Methods in Ecology and Evolution, 7*(5), 573- 579. doi: 10.1111/2041-210X.12512

Lempert, H. (1989). Animacy constraints on preschool children's acquisition of syntax. *Child Development*, 237-245. https://doi.org/10.2307/1131088

Lenth, R. (2019). *Emmeans: estimated marginal means, aka least-squares means.* R package version, 1(2).

Lieburg, R. van, Sijyeniyo, E., Hartsuiker, R.J., & Bernolet, S. (2022). *The development of abstract syntactic representations in beginning learners of Dutch.* Manuscript in preparation.

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language, 91.* https://doi.org/10.1016/j.jml.2016.03.009

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2005). The role of frequency in the acquisition of English word order. *Cognitive Development, 20*(1), 121-136. https://doi.org/10.1016/j.cogdev.2004.08.001

Messenger, K., & Fisher, C. (2018). Mistakes weren't made: Three-year-olds' comprehension of novel-verb passives provides evidence for early abstract syntax. *Cognition, 178,* 118-132. https://doi.org/10.1016/j.cognition.2018.05.002

Messenger, K., Branigan, H. P., McLean, J. F., & Sorace, A. (2012). Is young children's passive syntax semantically constrained? Evidence from syntactic priming. *Journal of Memory and Language, 66*(4), 568-587. https://doi.org/10.1016/j.jml.2012.03.008

Montgomery, J. W., Gillam, R. B., Evans, J. L., & Sergeev, A. V. (2017). "Whatdunit?" Sentence comprehension abilities of children with SLI: Sensitivity to word order in canonical and noncanonical structures. *Journal of Speech, Language, and Hearing Research, 60*(9), 2603-2618. doi: 10.1044/2017_JSLHR-L-17-0025

Muylle, M., Wegner, T. G., Bernolet, S., & Hartsuiker, R. J. (2020). English norming data for 423 short animated action movie clips. *Acta Psychologica, 202*, 102957. https://doi.org/10.1016/j.actpsy.2019.102957

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution, 4*(2), 133 142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language, 81*, 1-15. https://doi.org/10.1016/j.jml.2014.12.002

Pickering, M., & Ferreira, V. (2008). Structural Priming: A Critical Review. *Psychological bulletin, 134*, 427–459. https://doi.org/10.1037/0033-2909.134.3.427

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and language, 39*(4), 633-651. https://doi.org/10.1006/jmla.1998.2592

Powell, M. J. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, 26.*

R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of child language, 31*(1), 61 99. https://doi.org/10.1017/S0305000903005956

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly journal of experimental psychology, 67*(6), 1176-1190. https://doi.org/10.1080/17470218.2013.850521

Verrips, M. (1996). *Potatoes must Peel. The Acquisition of the Dutch Passive.* Ph.D. Dissertation, University of Amsterdam.

Appendix S1

Transitive Verbs Used in Picture Set.

Aanrijden, *bellen*, *dragen* and *volgen* occurred 20 times in the picture set. The rest of the verbs occurred 16 times in the picture set.

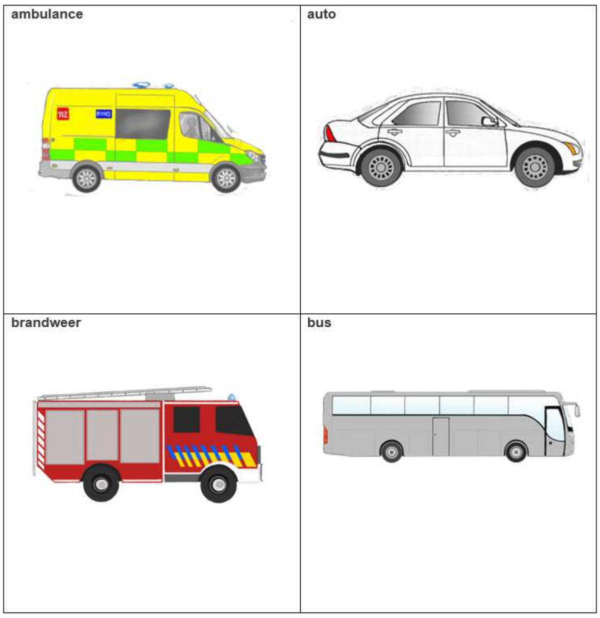
Dutch Verb	**English Translation**	**Zipf-frequency (Keuleers et al., 2010)**	**AoA (Brysbaert et al., 2014)**
aanrijden	to hit	3.25	9.28
bedienen	to serve	3.85	8.89
bellen	to call	5.35	6.03
blokkeren	to block	3.67	8.72
dragen	to carry	4.91	5.78
groeten	to greet	4.48	6.71
helpen	to help	5.76	5.40
inhalen	to overtake	2.72	8.61
roepen	to shout	4.53	4.96
slepen	to drag	3.95	8.03
vervangen	to replace	4.41	9.06
volgen	to follow	4.99	5.84

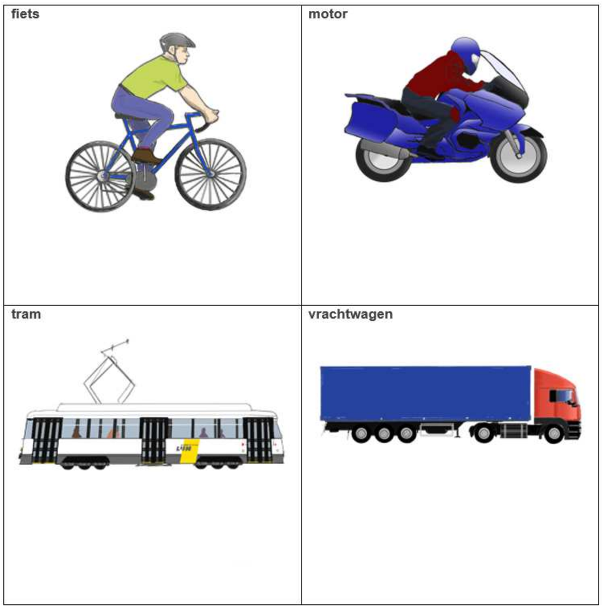
List of Human Professions and Vehicles

Human Professions	**English translation**	**Vehicles**	**English translation**
bakker	baker	ambulance	ambulance
bouwvakker	construction worker	auto	car
dokter	doctor	brandweer	fire truck
journalist	journalist	bus	bus
kok	cook	fiets	bike
leraar	teacher	motor	motor
slager	butcher	tram	tram
zangeres	singer	vrachtwagen	truck

Human professions

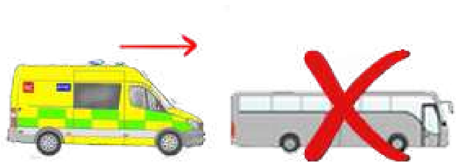
Vehicles


ambulance
auto
brandweer
bus

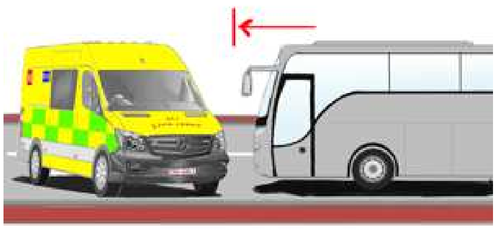

fiets
motor
tram
vrachtwagen

Appendix S2

Examples of pictures with the abstract verbs ‘vervangen’ [to replace]; ‘blokkeren’ [to block].



An example of a picture that depicts the sentence ‘De ambulance vervangt de bus’ – The ambulance replaces the bus



An example of a picture that depicts the sentence ‘De ambulance blokkeert de bus’ – The ambulance blocks the bus.

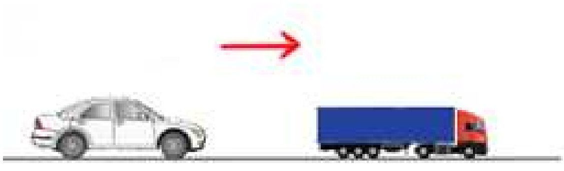
Appendix S3

Table 3: Full output of final model with the predictors Condition (active condition as reference level), Verb (aanrijden as reference level) and Grade (high grade as reference level).

Summary of the fixed effects in the multilevel logit model (N = 8278; **log-likelihood** = -1594.9) **Fixed effects**	β-coefficient	SE	Z-value	p-value
(Intercept)	-5.58	0.51	-10.89	<.001 ***
Condition (passive)	1.55	0.40	3.86	<.001 ***
Verb (bedienen)	-0.67	0.90	-0.74	0.45
Verb (bellen)	-1.57	0.88	-1.78	0.07 .
Verb (blokkeren)	1.21	0.67	1.80	0.07 .
Verb (dragen)	-0.79	0.82	-0.96	0.33
Verb (groeten)	-0.31	0.80	-0.38	0.69
Verb (helpen)	0.16	0.74	0.22	0.82
Verb (inhalen)	2.37	0.58	4.06	<.001 ***
Verb (roepen)	-2.06	1.16	-1.77	0.07 .
Verb (slepen)	0.17	0.75	0.22	0.81
Verb (vervangen)	2.69	0.58	4.62	<.001 ***
Verb (volgen)	-1.58	0.91	-1.73	0.08 .
Grade (low)	1.80	0.41	4.37	<.001 ***
Condition (passive) * Verb (bedienen)	-0.88	0.57	-1.54	0.12
Condition (passive) * Verb (bellen)	1.26	0.72	1.74	0.08 .
Condition (passive) * Verb (blokkeren)	-1.89	0.56	-3.37	<.001 ***
Condition (passive) * Verb (dragen)	-0.08	0.64	-0.12	0.90
Condition (passive) * Verb (groeten)	-0.62	0.54	-1.15	0.24
Condition (passive) * Verb (helpen)	0.21	0.64	0.32	0.74
Condition (passive) * Verb (inhalen)	-1.12	0.46	-2.41	<.05 *
Condition (passive) * Verb (roepen)	2.11	1.06	1.98	<.05 *
Condition (passive) * Verb (slepen)	-1.09	0.56	-1.93	0.05 .
Condition (passive) * Verb (vervangen)	-0.82	0.47	-1.74	0.08 .
Condition (passive) * Verb (volgen)	2.30	0.81	2.82	<.01 **
Verb (bedienen) * Grade (low)	-1.47	0.80	1.84	0.06 .
Verb (bellen) * Grade (low)	0.63	0.62	1.02	0.30
Verb (blokkeren) * Grade (low)	-0.07	0.58	-0.13	0.89
Verb (dragen) * Grade (low)	0.37	0.65	0.57	0.56
Verb (groeten) * Grade (low)	1.52	0.70	2.18	<.05 *
Verb (helpen) * Grade (low)	-0.75	0.55	-1.36	0.17
Verb (inhalen) * Grade (low)	0.38	0.45	0.85	0.39
Verb (roepen) * Grade (low)	0.19	0.59	0.31	0.75
Verb (slepen) * Grade (low)	0.53	0.64	0.83	0.40
Verb (vervangen) * Grade (low)	-0.75	0.45	-1.66	0.09 .
Verb (volgen) * Grade (low)	0.07	0.50	0.15	0.88

Appendix S4

The Transitive Picture for “The car follows the truck”. Note the use of the Arrow.



The arrow indicates the direction of the transitive event. It is likely that people will first process “the car” as the agent, leading to an active interpretation, before processing “the truck”. Confusion may arise when people hear “the truck” first, since the direction of the arrow should then be disregarded.

STEM-, SPRAAK- EN TAALPATHOLOGIE
Vol. 27, 2022, Special Issue, pp. 206-222
Special Issue for Steven Gillis
Gepubliceerd online: 5 november 2022

University of Groningen Press
https://doi.org/10.21827/32.8310/2022-SG-206

Yes or No: How children combine gestures and speech to express honest and deceiving attitude

Marc Swerts[1], Yan Gu [2,3], & Tarissa Boerrigter[1]

[1] *School of Humanities and Digital Sciences, Tilburg University*
[2] *Department of Experimental Psychology, University College London*
[3] *Department of Psychology, University of Essex*

Samenvatting

Dit artikel rapporteert over een studie naar de manier waarop kinderen hoofdbewegingen maken die bedoeld zijn om hun appreciatie voor een getoond object uit te drukken, waarbij we een vergelijking maken tussen gevallen waarbij die hoofdbeweging wel of niet overeenkomt met hun echte waardering. Aan vierenveertig kinderen tussen vijf en acht jaar oud werd gevraagd om aan een experimentator te vertellen of ze een reeks getoonde objecten wel of niet als cadeau voor hun verjaardag zouden willen krijgen. In een eerste ronde deden kinderen die taak zonder extra instructie zodat hun antwoord overeenkwam met hun echte waardering van een object (congruent). In een tweede ronde werd hen gevraagd om een antwoord te geven dat tegengesteld was aan hun echte waardering (incongruent). Analyses van hun verbale antwoorden en reactietijden laten zien dat jongere kinderen het moeilijker vonden om incongruente antwoorden te produceren. Naast het feit dat het relatieve aantal hoofdbewegingen afneemt met leeftijd blijkt dat kinderen uit alle onderzochte leeftijdscategorieën meer bewegingen maken in de congruente conditie, en meer verticale dan horizontale bewegingen.

Abstract

This study looks into children's use of head gestures to express their appreciation for objects, comparing cases in which the gestures match or do not match their true attitude. Forty-four children aged 5 to 8 years old were asked to tell an experimenter whether or not they would like to have shown objects as presents for their birthday. In a first round, children were not given any additional instructions, so that their feedback matched their genuine attitude towards the objects. In a second round, they were asked to give feedback in a way that was the opposite of what they felt. Analyses of their verbal reactions and response delays suggest that the youngest children found it harder to produce incongruent feedback. While the relative use of head gestures decreases with age, children in all age groups produce more head gestures in the congruent condition, and produce more shaking gestures.
Keywords: head gestures, signs of attitude, child development

Correspondentieadres:
Marc Swerts
Tilburg University
School of Humanities and Digital Sciences
Warandelaan 2, 5037 AB Tilburg
The Netherlands
E-mail: m.g.j.swerts@tilburguniversity.edu

Introduction

Children, as part of their communicative development, not only acquire the phonology, words and syntax of a language, but also learn how to interact via nonverbal features. One particular case at hand is the use of head movements: speakers exploit such gestures to support their spoken interactions, while their addressees derive meaningful information from them (e.g. Maynard 1987; McClave 2000). Darwin already remarked in the 19th century that many cultures use head gestures as cues to confirm or disconfirm information, and the most commonly found pattern is one where people make a succession of vertical movements (nodding) as an affirmative feedback signal to their conversation partners, whereas a series of horizontal movements (shaking) is used to deny information. This association of vertical movements with positive and horizontal movements with negative feedback has been argued by Darwin to have a natural origin: he hypothesized that this specific use of gestures and their interpretation stem from early childhood, where nodding is produced by babies when they search for milk, whereas a head shake comes from refusing the milk (Darwin 1872). At later stages in people's lives, the gestures get a more generic, or more symbolic, function of acceptance or refusal, and become used in a wider variety of contexts, not necessarily related to food.

The head gestures introduced above are often viewed as a subset of so-called "emblems", sometimes also termed "autonomous gestures" or "emblematic gestures". Emblems are defined as conventionalized bodily movements of hand, arm or head used by members of the same community to convey specific meanings which can even be understood without accompanying speech (Kendon 2004). A typical example of such an emblem is the use of a straight index finger that is vertically positioned on a person's closed lips to indicate that one should be quiet or silent. While some of these gestures appear to be used universally, others are less wide-spread, and it is even possible that seemingly identical gestures can serve different functions in different communities. For instance, the "thumbs up" or the "ring" gestures have a positive meaning in many western societies, i.e., they signal the equivalent of "okay" or "good" for residents of the United States, but they can have negative connotations in other cultures (Knapp & Hall 2010; Kita 2009; Matsumoto & Hwang 2013).

Similarly, regarding head gestures, it turns out that some cultures display a pattern which is the reverse of the more general trend, such as in Bulgaria where horizontal movements are being interpreted as "yes" and vertical movements as "no" (Andonova & Taylor 2012). The fact that the conventions about head movements may vary between communities and that their associated meanings can be opaque to members of a different culture suggests that their symbolic function needs to be acquired by children. That is, they cannot (or no longer) simply be derived from some intrinsic and natural form-function correspondence between a bodily movement and what this movement is supposed to depict, like is the case with a throwing gesture that people use to illustrate that specific act. One relevant question therefore is when and how children learn to use and interpret such emblematic gestures. In addition, analogous to what we know about other forms of linguistic or cognitive development, it is also interesting to explore how such gestures develop in growing children, i.e., whether people –with age- change in how they use emblems, just as they evolve regarding their lexical vocabulary and their use of syntactic structures. In that respect, there are a few developmental studies that focused on the acquisition of head gestures, even when the results of these studies are somewhat at variance.

An early study by Dittmann (1972) dealt with what he called listener responses, which included head nods (as well as a number of other features such as verbal acknowledgments like "yeah") that signal that an addressee is paying attention to what an interlocutor is saying. Interestingly, he found that children produce such responses to a far lesser degree than adults. It turned out that listeners responses were nearly absent in speech data of children in grades 1, 3 and 5, unless they were very explicitly urged by their conversation partner to signal some feedback. The relative frequency of such responses increases enormously by the time children are in grade 8, and becomes more adult-like by early adolescence. This pattern is reminiscent of claims regarding a more general development that a growing child changes from a largely egocentric individual to a person who is more socially aware of others (Saarni 1984; Swerts 2012). In other words, the work by Dittmann suggests that the use of head nods (i.e., as listener responses) becomes more frequent as a function of age, which would be related to an increased social awareness in a growing child.

However, more recent work suggests that developmental patterns in overt head gestures could be somewhat different, though those studies analysed a wider range of gestures (not only related to listener responses). Guidetti (2005) investigated French children at age 1;4, 2;0 and 3;0 while they were spontaneously interacting with their mothers, and specifically looked into refusal and acceptance messages which could be gestural, verbal or combinations of gestural and verbal. The study revealed that children in all age groups used nodding and shaking, often in combination with verbal messages, but the proportion of gesture-only messages turned out to be higher for younger children. Also, the study showed that assertions were more frequently expressed than refusals. Similarly, Fusaro, Harris and Pan (2012) studied child-mother dyads and observed children at 14, 20 and 32 months. They found that the frequency of these gestures increases with age, and that there is especially a rapid increase in nodding between 20 and 32 months, however mostly combined with speech. Additionally, children as they grow older more frequently begin to use such gestures in combination with speech, though isolated head nods continue to be used predominantly at 32 months. So both studies by Guidetti (2005) and Fusaro, Harris and Pan (2012) suggest that the absolute frequency of head gesture increases with age, but that the gestures become relatively less important, since proportionally speaking older children signal similar kinds of feedback more often with verbal messages.

In sum: developmental studies have brought to light that children fairly early in their life learn the "meaning" of nods and shakes, and also use them accordingly. However, the experimental studies by Guidetti (2005) and by Fusaro, Harris and Pan (2012) concentrated on fairly young children, so that it remains to be seen how a child's gestural behavior further develops at somewhat older ages. Also, to the best of our knowledge, child-directed research in this area has exclusively looked at contexts in which the child reactions in interactions with others are truthful, so that the horizontal and vertical head gestures are congruent with positive or negative messages that children intend to convey. To gain insight into the degree of automaticity with which such gestures are produced, the current study compares children's gestural behavior in truthful situations with situations where children are asked to express an attitude which is opposite to what they truly feel. We deal with Dutch children between the age of 5 and 8. Also, note that Dutch follows the general trend displayed in most cultures where nodding and shaking serve as positive or negative feedback signals, respectively.

Current study

Our current study has a number of goals. First, from the discussion of previous work it has become clear that typically-developing children fairly early during their life learn to associate head movements with symbolic meanings, and use them accordingly. So far, most child-directed studies of head gestures have focused on relatively young children and report that, with age, these children become more experienced users of such gestures. Not much is known, though, about developmental patterns in children who are somewhat older. From the literature, it is hard to predict what to expect regarding the development of head gestures in somewhat older children. On the one hand, one could argue that this relative use of head gestures will probably continue to increase, because the developing child becomes a more advanced user of these gestures, and becomes more knowledgeable about how to produce the gestures in a wider variety of situations. On the other hand, it has been shown that children between age 7 and 12, as they grow older, also become less expressive in terms of their nonverbal behavior (Shahid, Krahmer & Swerts 2008; Swerts 2012), even when this has mostly been shown for expressions of emotions. That is, older children tend to internalize their emotions more, and are better in controlling them and adapting them to specific social contexts. Based on this, and given that older children also have better developed verbal skills which could serve similar purposes as the nonverbal messages, it could be that children when growing older use head gestures less frequently.

Second, results from earlier work suggest that children exploit such head gestures almost automatically and spontaneously when they want to convey affirmative or negative feedback, especially given that earlier work focused on signals that are in line with the children's positive or negative attitudes. So when children are allowed to express their "true" internal state, this is very naturally and relatively frequently reflected in their head movements. That naturally begs the question as to whether such head gestures would still occur in cognitively more demanding contexts, such as when they would have to be deceptive about their attitude. Previous work has shown that lying is a more mentally challenging task for people than simple telling the truth (Vrij, Fisher, Mann & Leal 2006; Gombos 2006). This has, for instance, been shown by experimental work in which adult participants were invited to give answers to a series of questions in either a truthful or deceptive manner, and then appeared to be faster in the former task (Seymour, Seifert, Shafto & Mosmann 2000; Walczyk, Mahoney, Doverspike & Griffith-Ross 2009; Walczyk, Roper, Seemann & Humphrey 2003). Because of the additional mental processing tasks, children may typically behave in a less instinctive manner when they have to express attitudes which are in conflict with what they "really" feel. In a deceptive context, where there is a mismatch between the true attitude and what the head gesture displays, one could expect that the automaticity of gestures may get lost.

And third, developmental patterns in head gestures in growing children may be different for affirmative and negative head gestures. As we saw above, previous work on younger children has shown that, for them, acceptance and refusal gestures are quite distinct, in the sense that these children feel more inclined to signal affirmative than negative feedback (Guidetti 2005; Fusaro, Harris & Pan 2012). This may suggest that the latter type of gestures represents more marked behavior, even though it remains to be seen whether this also holds for older children. In addition, there is also a related, but more general finding that denying information is a cognitively more demanding activity than accepting information. From the point of view of comprehension, it has already been shown a few decades ago (Clark & Chase 1972; Wason 1959), as well as in many follow-up studies, that comprehenders find it harder to process sentences in which information is negated than sentences that contain affirmative information. For instance, if people are asked to verify whether the content of a sentence matches what is shown in a picture, then people are faster to respond to this task when the sentence is phrased in a positive (*X is above Y*) than in a negative manner (*X is not below Y*) (Clark & Chase 1972). While not going into the details of the proposed cognitive models that explain such results, this difference in verification latency has been interpreted as evidence that positive sentences represent a "simpler" code. So, even though previous studies were mostly perceptual/comprehensive in nature, and were not concerned with head gestures at all, they do suggest that affirmative and negative actions are mentally quite distinct, which may have repercussions on how participants display different kinds of feedback in a production experiment.

Given the three goals of this study, in the following sections, we first describe our method to elicit truthful and deceptive signs of appreciation in children of different age groups, and then we present and discuss the results of various analyses, both on overall behavior (e.g., verbal responses; reaction time) in those various contexts, and more specifically on differences in head gestures.

Method

Participants

44 children (20 girls) from the same school participated in the experiment. The mean age of children was 6 years and 6 months. These children were distributed in groups 2, 3 and 4 of their primary school in the following way: 10 (7 girls) in group 2, 20 (8 girls) in group 3 and 14 (5 girls) in group 4, with average ages of 5 years and 2 months, 6 years and 3 months, and 7 years and 2 months, respectively. We decided not to include younger children (group 1), as we expected the task to be too difficult for them. Children were allowed to participate only if the experimenters had received an informed consent from their parents and school. All children were given a small gift as a token of appreciation. Data from one child from the youngest group were not further analysed because that child did not manage to do the "reverse world" condition (see next session).

Procedure

Children were asked to inform an experimenter (the third author of this article) as to whether they liked or disliked objects that were presented to them as a series of pictures on a computer screen of a laptop positioned in front of them. While the experimenter was physically present in the same room, the children could not see her. In particular, they were instructed to signal whether they would love a specific object (like a swimming pool) as a birthday present or not. In order to make sure that this set-up would lead to both affirmative and negative reactions from children, we selected objects which presumably would be liked by quite a number of children, and presents (e.g. an onion) that would probably lead to opposite reactions. The selection of the objects was based on some pretesting in which we checked whether pictures could easily be recognized, and we tried to design the list in such a way that we would likely get about an equal number of positive and negative responses from children. Also, we designed two lists that varied a bit, depending on whether the pictures would be shown to boys or girls, such that we, for instance, used a doll in the list for the girls, and a football for the boys. In this way, twenty objects in total were randomly presented to participating children in 2 conditions (see below). It is important to note that children were only instructed to inform the experimenter about their appreciation of the objects, and were not explicitly asked to use head gestures.

The procedure was such that we first elicited responses from children (in an individually performed experiment) in a truthful manner (congruent condition), whereby we did not present children with any specific instructions to them other than that they had to indicate whether or not they would appreciate a specific object as a present or not. After they had done this test, they were again presented with the same list of objects, but this time they were asked to do the task under a "reverse world" condition, which was explained to them as a condition in which they had to express the opposite of what they thought (incongruent condition). Both conditions were preceded by a short practice session, in which the children in the presence of the experimenter would give a reaction to some trial objects that were not used in the actual experiment later on. During this practice session, the experimenter did not actively participate herself (so did not respond to the stimuli), but was simply there to check whether the child understood the task. After this practice session, the experimenter would disappear out of sight (to avoid that children would search for eye contact or would get influenced by the experimenter's behavior) but stayed in the same room, and the children would then do the actual experiment. While this set-up was chosen to ensure experimental control, the disadvantage is of course that the child is not communicating anymore with a real person in a face-to-face setting. (We elaborate on this issue in the general discussion.) The experimenter would remain silent during the actual experiment, though she could see and hear the participating children. It turned out that both tasks could easily be understood by the children, who also informed the experimenter afterwards that they thought it was a fun experiment. The whole procedure, including the practice sessions, took about 6 minutes on average per child.

The objects were presented as pictures to the children in a powerpoint presentation, the speed of which was paced by the experimenter via a remote control. She always showed a new slide immediately after a child had responded to a preceding slide. To facilitate the measurements of reaction times afterwards, a new picture was always presented simultaneously with a specific nonverbal computer sound. The presentation was shown on an Apple Macbook Pro versie 10.7.4. During the experiment, the participating children were video-recorded with a built-in FaceTime HD camera that had a 1280 by 720 resolution, so that children's responses and head gestures could be analysed at a later stage. Video recordings were saved as iMovie clips, and the audio streams were extracted from these recordings to enable the measurement of response delays in Praat.

Annotation and measures

To better understand how adequately children could do the task to confirm or disconfirm information in a truthful or deceptive manner, and how quickly they were able to respond under various settings we coded the following aspects:
Responses: Children's affirmative and negative feedback cues were coded for each trial. In the incongruent (deceptive) condition, if a child failed to reverse the feedback, it was coded as an error. For instance, when a child in the congruent condition stated that it did not like a potato as a birthday present, we classified an identical reaction in the incongruent condition as an error.
Reaction time: Given the earlier findings by Walczyk and colleagues (Walczyk, Mahoney, Doverspike & Griffith-Ross 2009; Walczyk, Roper, Seemann & Humphrey 2003) that response time is a good indicator of whether people are being truthful or deceptive, we measured how long children took to produce an utterance after the picture had been shown (using the computer sound presented in the powerpoint as a reference). The large majority of the participants' responses were produced with speech, except for 4 children from the youngest group who regularly or always responded with a nonverbal head gesture only, so these "silent" responses were not taken into account in the current measurement, also given their infrequent occurrence.
Head gestures: the third author of this article annotated all the responses of the children in terms of presence or absence of a gesture, and in case of presence, determined whether a gesture was of a vertical (nodding) or horizontal (shaking) type and whether the gesture matched the required responses. A matching gesture was defined as a head movement that corresponded with the affirmative or negative status of the feedback (affirmative: nodding; negative: shaking). We did not perform detailed analyses of how the non-verbal gestures were aligned with the verbal responses.

Data Analysis

We used logistic regression (glmer in R environment, R Core Team, 2020) to analyze binary dependent variables (responses errors; proportion of nodding/shaking gestures, overall gesture use, matching gestures) and linear mixed effects models to analyze reaction times. Reaction times were log-transformed, following Whelan (2008).

Children's age, valence (affirmative, negative) and experimental condition (congruent, incongruent) were included as fixed effects, as well as their two-way and three-way interactions. To account for the individual differences of participants and item differences of target words, we included participants and stimuli as two grouping variables for the random effects structure of the models. For both participants and words, the structure consisted of a random intercept but not necessarily a random slope of condition or valence due to the model convergence.

When the model did not converge, we ran the models with different optimizers that allow convergence or reduced the model complexity or mean-centering independent variables before analyses. Data files in .csv format and all analysis scripts and R markdown output can be found at the OSF link: https://osf.io/vbwgx/.

Results

Overall distribution of responses

Overall, it turns out that children produced about 820 affirmative and 940 negative feedback cues. A binomial test indicated that the proportion of 0.466 for affirmative feedback was significantly lower than the expected proportion of 0.534 for negative feedback, $p = .0001$ (1-sided). The proportion of affirmative feedback was not significantly different over ages (β = -0.062, p=.29).

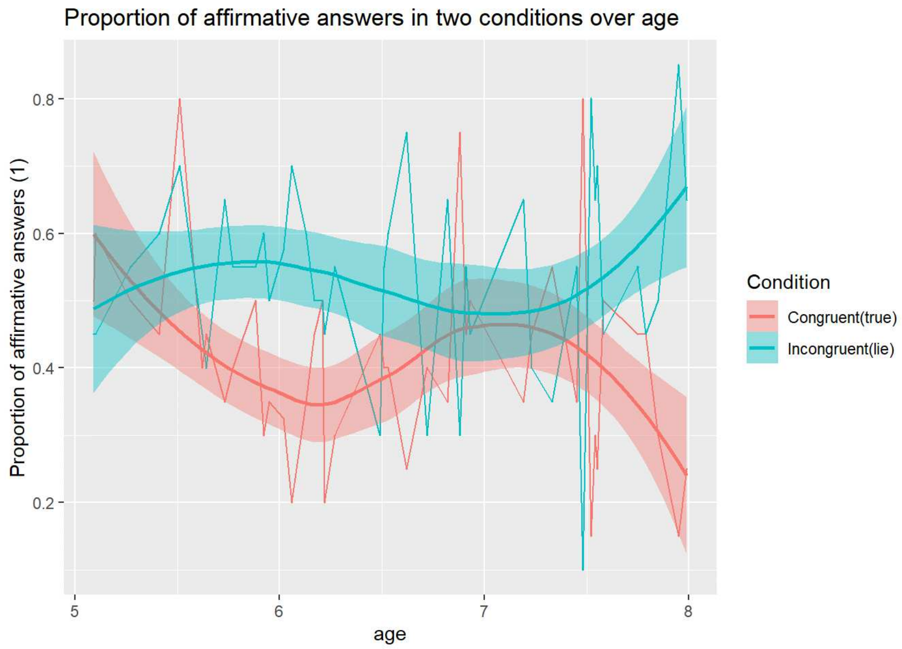


Figure 1: Affirmative answers in congruent and incongruent (lie) conditions over age.

In the "ideal" case where children would always manage to be deceptive about their true attitude, a proportion of an x% of affirmative (negative) reactions in the incongruent condition should present a proportion of 1-x% of these reactions in the congruent condition, given that the former condition should literally reveal the opposite pattern of the latter one. Fig. 1 illustrates children's affirmative answers in congruent and incongruent (lie) conditions at different ages. As can be seen, children of 5 or younger than 6.5 years may still find it more difficult to be deceptive about their appreciation, thus making "errors" (a response to an object in the incongruent condition turned out not to be the exact opposite of what the child had responded in the congruent condition).

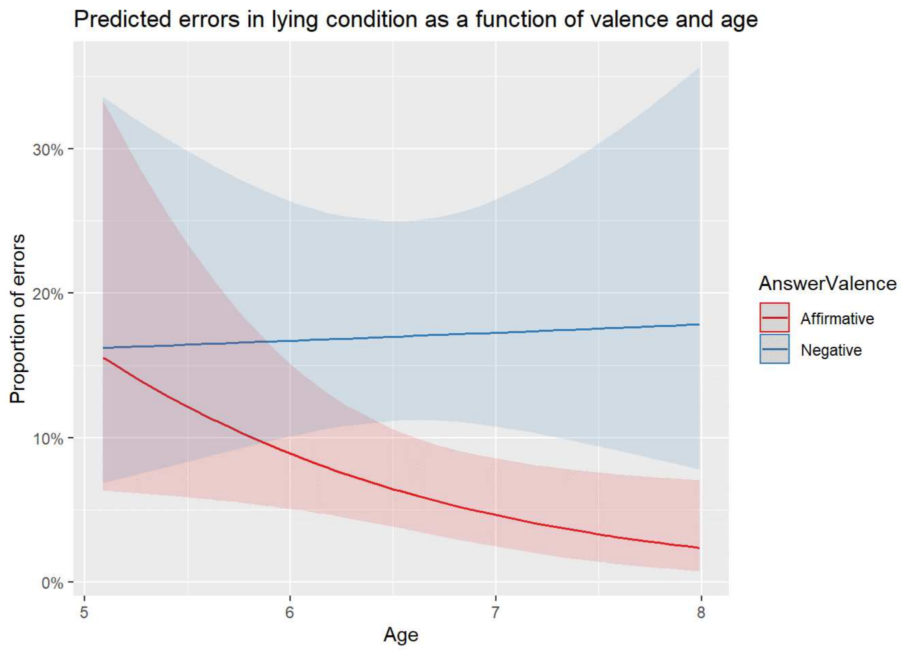


Figure 2: Percentage of errors in the incongruent condition as a function of response valence and age.

To examine this age effect, we studied whether the percentages of "errors" in the incongruent condition change as a function of children's ages and responses valence (affirmative and negative responses). The results of a logistic regression showed that children made more errors when they deceive about negative valence than affirmative (β = 1.24, p < .001). That means it was harder to hide negative valence. Age was not significant (β = -0.24, p =.38), but there was an interaction between age and valence of responses (β = 0.73, p = .006). The interaction term revealed that an increase in age related to fewer errors in lying about affirmative valence (β = -.70, p =.031) but did not relate to fewer errors in lying about negative valence (β = .04, p = .89). This can also be seen in Fig. 2: while the youngest children made about an equal amount of errors for both the affirmative and negative actions, the errors appeared to be relatively more frequent for negative responses in the older children.

Reaction times

Table 1 presents the reaction times in seconds (means and standard error) for affirmative and negative responses as a function of age group and experimental condition. The analysis of a linear mixed-effects model revealed a main effect of condition (β = .084, p < .001), while the effect of age (β = -.02, p = .99) and valence (affirmative vs negative) (β = -.0054, p = .55), and their three-way interactions were all not significant (β = .014, p = .51). As shown in Figure 3, incongruent feedback was produced after a significantly longer delay than congruent. However, there was an interaction between age and condition (β = -.047, p < .001), such that the differences between conditions were larger in younger than older children (β = .019, p = .02).

Table 1: Reaction times in seconds (means and standard error) for affirmative and negative responses as a function of age group and experimental condition

		Response	
Group	Condition	Affirmative	Negative
2	Congruent	1.73 (.20)	1.65 (.14)
	Incongruent	1.95 (.21)	2.21 (.23)
3	Congruent	1.12 (.12)	1.31 (.09)
	Incongruent	1.77 (.13)	1.48 (.14)
4	Congruent	1.43 (.14)	1.40 (.10)
	Incongruent	1.61 (.15)	1.56 (.16)

Gestures

Overall, children produced 223 shaking and 107 nodding head gestures. A binomial test indicated that the proportion of shaking of 67.58% was significantly higher than the expected nodding of 32.42%, p < .0001 (1-sided). Results of a logistic regression analysis showed that the proportion of noddings decreased (when compared to shakings) with increasing age (β = -.37, p = .057). There was a marginally significant effect of condition (β = .48, p = .085), resulting from the fact that the incongruent condition had a higher proportion of noddings (Figure 4).

Furthermore, we examined the overal use of head gestures (nodding and shaking) as a function of condition, response valence and age. The analysis revealed a main effect of valence (β = 1.41, p < .001) (negative responses lead to more gestures than affirmative; negative: 21.4%; positive: 12.5%), and marginally significant effects of condition (β = -1.49, p = .069) (congruent condition leads to more gestures than incongruent; congruent: 22.0%; incongruent: 12.4%) and age (β = -1.32, p = .054). There was a two-way interaction between answer valence and condition (β = -1.30, p = .015), indicating that for the negative valence

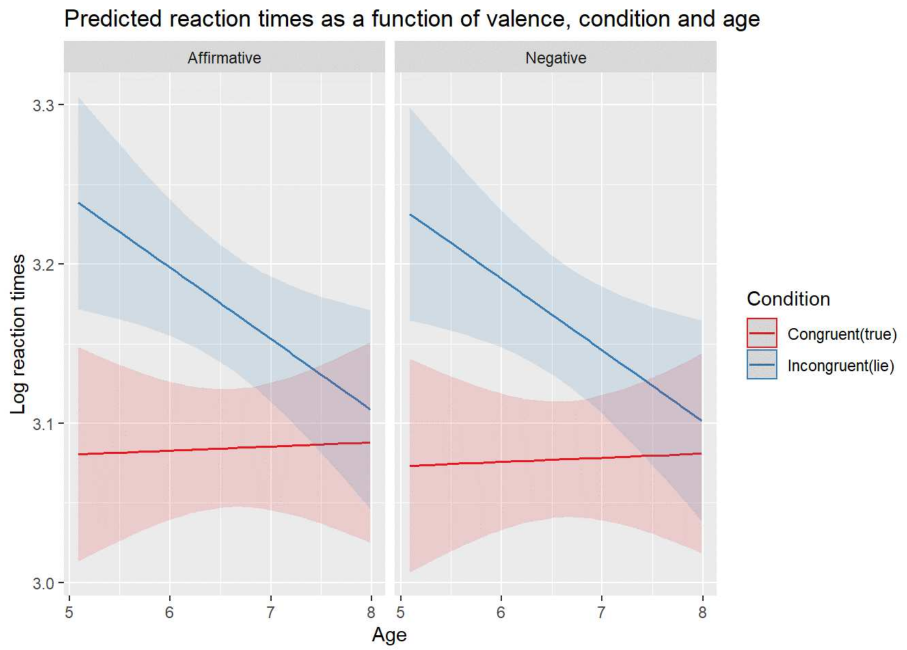


Figure 3: Predicted reaction times (log-transformed) as a function of response valence, condition and age.

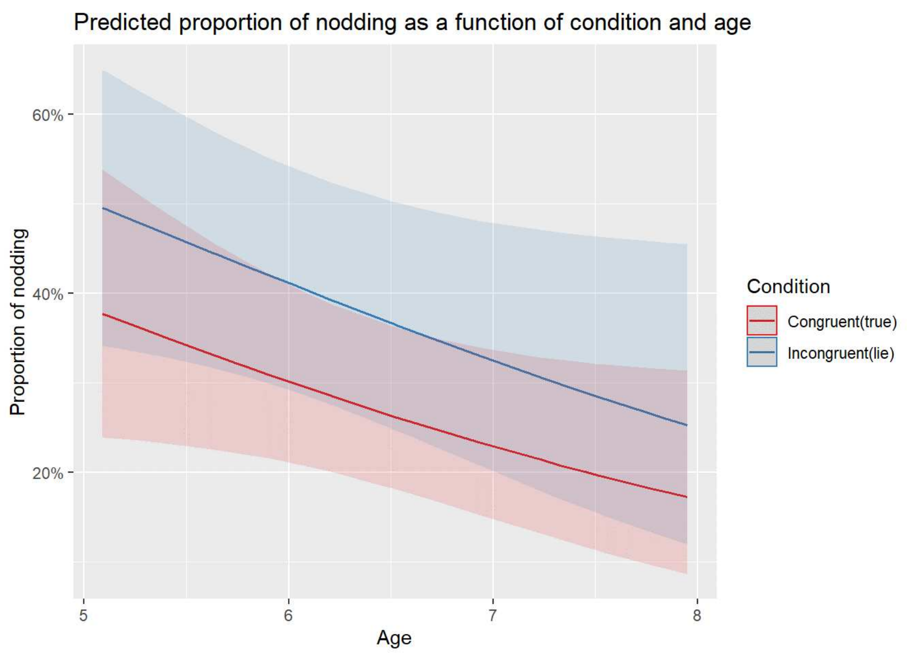


Figure 4: Predicted percentages of noddings as a function of response valence and age.

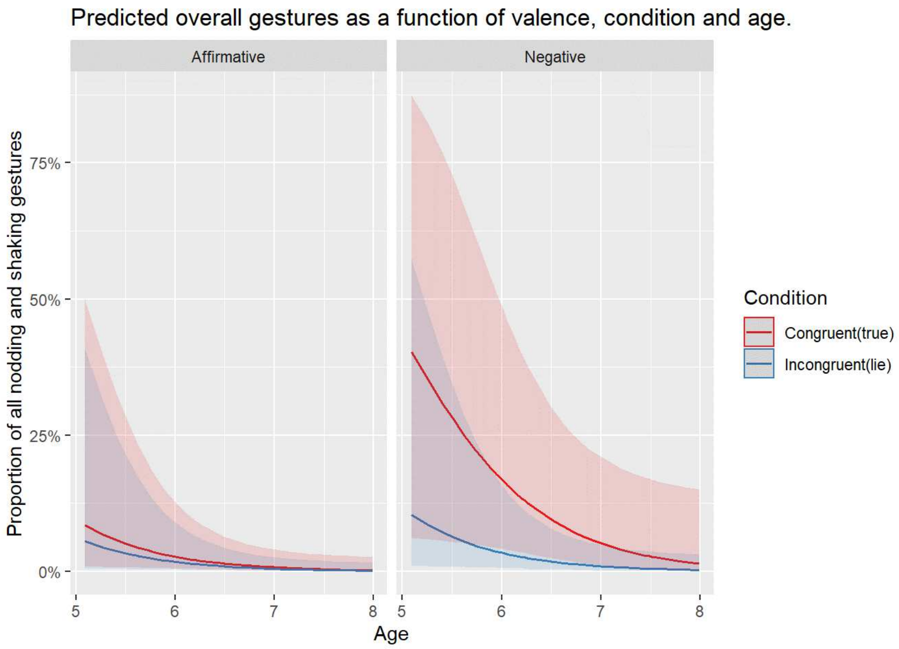


Figure 5: Predicted percentages of the overall use of nodding and shaking gestures as a function of response valence, condition and age.

congruent condition had more gestures than the incongruent condition ($\beta = 1.99$, $p < .001$) (see Fig. 5). There were no other two-way or three-way interactions.

The results were quite similar when we examined the proportion of matched head gestures as a function of valence, age and condition. There were main effects of response valence ($\beta = 1.48$, $p <.001$), marginally significant effects of condition ($\beta = -1.10$, $p = .0495$), age ($\beta = -1.20$, $p = .0755$), and an interaction between valence and condition ($\beta = -1.12$, $p = .04$).

Finally, there were 114 responses that were only produced head gestures. Analysis showed that there were main effects of age ($\beta = -5.53$, $p =.011$), condition ($\beta = 14.06$, $p < .001$) and an interaction between age and condition ($\beta = 16.18$, $p < .001$). It indicated that gesture-only responses declined when children were older but this only appeared in the congruent condition ($p = .0004$).

General discussion and conclusion

The present study has presented an analysis of verbal answers and gesture responses (nods and shakes) produced by children in different ages, when they were asked to indicate their appreciation for various objects that were shown to them as a series of pictures. There were two experimental conditions: a congruent one in which children could express their true attitude, and an incongruent one, in which children were asked to signal the opposite of what they really felt. Overall, we found that the latter condition appeared to be the more "demanding" condition, and that the experienced difficulty of the task decreased as a function of age. Overall, younger children made more errors than older ones when they had to give their appreciative feedback in the incongruent condition. In addition, while the proportion of errors was about the same for affirmative and negative responses in the youngest group, the older groups made relatively more errors for the negative feedback cases. Also, in line with earlier work, it turned out that children overall took significantly longer to respond to a shown picture in the incongruent than in the congruent condition. However, we found that such differences were smaller for older children.

When we look at their actual gestural behavior, it was found that the relative use of gestures depends on whether they are used in a congruent or incongruent manner. More specifically, we found that gestures are more frequently used in the congruent condition, and the overall number of gestures in the congruent condition decreases as a function of age (see below). Possibly, this result is related to the fact that the incongruent condition is cognitively demanding, as also became clear from the results of the response latencies above. In general, it has been found that people find it more difficult to express nonverbal expressions when these are not consistent with the internal states of a speaker. In a study on the expression of emotions, for instance, Wilting, Krahmer and Swerts (2006) found that spontaneous expressions of negative or positive feelings appear more natural and less marked, compared to acted versions of these emotions that are not "felt" as such by the speakers. Along the same lines, it may have been harder for the children to produce head gestures when these are not in line with their true attitude towards an object. Other studies on deception also revealed that producing a lie may have consequences for a person's nonverbal behavior, even when studies vary regarding the extent to which alleged nonverbal correlates of deception are consistent and reliable (DePaulo, Lindsay, Malone, Muhlenbruck, Charlton & Cooper 2003). In our own previous work on deception in children (Swerts 2012; Swerts, Verhoofstad & van Doorenmalen 2013), where we used a computerized puppet show to elicit minimal pairs of children's truthful and deceptive utterances, we found that observers can detect the deceptive utterances (in pairwise comparisons) from more marked settings of specific facial characteristics, even when these results were only slightly above chance level. Similarly, the relative absence of spontaneous head gestures in the deceptive condition of our current study could be due to the fact that the mentally demanding task had an inhibitory effect on the children's natural expressive behavior in terms of head gestures.

These results about differences between congruent and incongruent settings on head gestures are also in line with earlier findings on cognitive, "natural" associations between overt head gestures and negative or positive connotations. Overt head gestures are not only relevant for an interlocutor who observes these gestures from his/her conversation partner as feedback cues about the ongoing interaction, but, in addition, such gestures may also positively or negatively affect the attitude of the person who is producing them. Experimental studies in which participants were invited to either nod (affirmative) or shake (negative) their heads while doing specific tasks showed that the former type of movements led these participants to become more positive about a product or an idea, while the reverse appeared to be true with the latter type of gestures (Wells & Petty 1980; Briñol & Petty 2003). These results are consistent with a self-validating hypothesis which predicts that gestures have a biasing effect that can enhance or inhibit specific attitudes of people. The current study analyses participants' behavior when this natural link is being disrupted, i.e., when people are asked to express their appreciation for objects in a deceptive manner, so signaling the opposite of what their real attitude is, this less likely leads to the use of head gestures.

Next, it appears that there is an overall decrease in the use of head gestures and gesture only responses as a function of age. This is reminiscent of earlier work on children, though those studies focused on younger children than the ones investigated in the current study (e.g. Guidetti 2005). It is also in line with earlier work that showed that children display fewer nonverbal correlates of their emotions as they grow older (e.g. Shahid, Krahmer & Swerts 2008). The difference in the relative frequency of head gestures between younger and older kids could be related to the fact that the verbal skills of children develop with age. Young children learn fairly early what the symbolic, emblematic meaning is of head gestures, especially as these have been argued to have a natural origin (Darwin 1872). While younger kids may over-use nonverbal gestures to communicate particular things, the older children gradually rely more on verbal language for their messages, and exploit head gestures in a more controlled manner so that they suit specific social contexts.

And finally, if we compare the nodding and shaking gestures shaking gestures are significantly more frequently used than nodding gestures, and this appears to be true for every age group that we analysed in this study. At first sight, this seems to be in conflict with earlier studies (Guidetti 2005; Fusaro, Harris & Pan 2012) with their reported findings that nodding occurs more frequently than shaking. However, this appears to be true only when we consider absolute frequencies, where it was indeed found that children more often signal acceptance than refusal. Looking more closely at the data by Guidetti (2005), however, reveals that the proportion of those acts that contain a gesture (either nodding or shaking) is actually higher for the refusals, which finding is in line with what we have found. In other words, while it may be true that positive feedback signals predominate in natural interactions of young children, the relative usage of head gestures to support such acts is higher for refusals. One possible explanation for this is that children feel a higher need to signal denials by means of (additional) gestures, because these kinds of feedback cues represent more marked cases. In that sense, the results are in line with earlier work on positive vs negative feedback cues that showed that the latter type tend to be produced with more prominent prosodic features (e.g. higher pitch, slower tempo, longer delay), as it is often more crucial for communication partners to detect negative feedback than positive feedback because they signal a (potential) problem in the ongoing interaction (Shimojima, Katagiri, Koiso & Swerts 2002).

Of course, there are different ways in which this research could be extended. In the future, it could be useful to investigate in more detail why there are more noddings in the incongruent condition than the congruent condition, and whether possible differences in gesture behavior could provide cues to observers as to whether a person is truthful or deceptive. Future work could also focus on more interactive contexts. Basically, even when the experimenter in the study described above was present in the same room (but silent and not visible to the child participant), the children were basically talking to the computer.

It is therefore interesting to see what would happen if the children would have to do a similar task where they have to convince a human partner. One reason why this situation could lead to different, possibly more marked expressions, is that the mere presence of a human dialogue partner may make the task more demanding, or may make the children more aware of their deceptive act, which has been shown to have repercussions for a person's nonverbal behavior (Vrij, Fisher, Mann, & Leal 2006; Swerts, van Doorenmalen & Verhoofstad 2013).

Acknowledgments

We thank the children and their parents, as well as the teachers and director of the Trinoom elementary school in Eindhoven (The Netherlands) for their willingness to participate in the experiment reported upon in this article. Parts of the research of this article have been presented earlier at the 36th Annual Cognitive Science Conference (July 2014, Quebec City, Canada).

Literature

Andonova, Elena & Taylor, Holly.A. (2012). Nodding in dis/agreement: a tale of two cultures. *Cognitive Processing 13*, 79-82.

Briñol, Pablo & Petty, Richard E. (2003). Overt head movements and persuasion: a self-validation analysis. *Journal of Personality and Social Psychology 84* (6), 1123-1139.

Clark, Herbert H. & Chase, William G. (1972). On the process of comparing sentences against pictures *Cognitive Psychology 3*, 472-517.

Darwin, Charles (1872) *The expression of the emotions in man and animals* (3rd edition). Oxford University Press [1998].

DePaulo, Bella M. (1988). Nonverbal aspects of deception. *Journal of Nonverbal Behavior 12*, 153–161.

DePaulo, Bella M., Lindsay, James J., Malone, Brian E., Muhlenbruck, Laura, Charlton, Kelly, & Cooper, Harris (2003). Cues to deception. *Psychological Bulletin 129*, 74–118.

Dittmann, Allen T. (1972). Developmental factors in conversational behavior. *Journal of Communication 22* (4), 404-423.

Fusaro, Maria, Harris, Paul L. & Pan, Barbara A. (2012). Head nodding and head shaking gestures in children's early communication. *First Language 32* (4), 439-458.

Gombos, Victor A. (2006). The Cognition of Deception: The Role of Executive Processes in Producing Lies. *Genetic, Social, and General Psychology Monographs 132* (3), 197-214.

Guidetti, Michèle (2005). Yes or no? How young French children combine gestures and speech to agree and refuse. *Journal of Child Language 32*, 911-924.

Kendon, Adam (2004). *Gesture: Visible Action as Utterance.* Cambridge, UK: Cambridge University Press.

Kita, Sotaro (2009). Cross-cultural variation of speech-accompanying gesture: a review. *Language and Cognitive Processes 24* (2), 145-167.

Knapp, Mark L. & Hall, Judith A. (2010). *Nonverbal Communication in Human Interaction* [7th edition]. Wadsworth: Cengage Learning.

Matsumoto, David & Hwang, Hyi Sung (2013). Cultural similarities and differences in emblematic gestures. *Journal of Nonverbal Behavior 37*, 1-27.

Maynard, Senko K. (1987). Interactional functions of a nonverbal sign: Head movement in Japanese dyadic casual conversation. *Journal of Pragmatics, 11*, 589–606.

McClave, Evelyn Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics, 32*, 855–878.

R Core Team (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

Saarni, Carolyn (1984). An observational study of children's attempts to monitor their expressive behavior. *Child Development 55*, 1504-1513.

Seymour, Travis L., Seifert, Colleen M., Shafto, Michael G. & Mosmann, Andrea L. (2000). Using response time measures to access "Guilty Knowledge". *Journal of Applied Psychology 85* (1), 30-37.

Shahid, Suleman, Krahmer, Emiel J., & Swerts, Marc (2008). Alone or together: Exploring the effect of physical co-presence on the emotional expressions of game playing children across cultures. In P. Markopoulus (Ed.), *Fun and Games* (pp. 94-105). Berlin/Heidelberg: Springer. (Lecture Notes in Computer Science LNCS, 5294).

Shimojima, Atsushi, Katagiri, Yasuhiro, Koiso, Hanai & Swerts, Marc (2002). Informational and Dialogue-Coordinating Functions of Prosodic Features of Japanese Echoic Responses. *Speech Communication 36*, 113-132.

Swerts, Marc (2012). Correlates of social awareness in the visual prosody of growing children. *Laboratory Phonology 2* (2), 381-402.

Swerts, Marc, van Doorenmalen, Anniek & Verhoofstad, Lynn (2013). Detecting cues to deception from children's facial expressions: On the effectiveness of two visual manipulation techniques. *Journal of Phonetics 41* (2013), pp. 359-368.

Vrij, Aldert, Fisher, Ronald, Mann, Samantha, & Leal, Sharon (2006) Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences 10*, 141–142.

Walczyk, J.J., Mahoney, KT., Doverspike, D. & Griffith-Ross, D.A. (2009). Cognitive lie detection: Response time and consistency of answers as cues to deception. *Journal of Business and Psychology 24* (1), 33-49.

Walczyk, Jeffrey J., Roper, Karen S., Seeman, Eric, & Humphrey, Angela M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology 17*, 755–774.

Wason, Peter C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology 11* (2), 92-107.

Wells, Gary L., & Petty, Richard E. (1980). The effects of overt head movements on persuasion: Compatibility and incompatibility of responses. *Basic and Applied Social Psychology 1*, 219–230.

Whelan, Robert (2008). Effective analyses of reaction time data. *The Psychological Record 58*, 475-482.

Wilting, Janneke, Krahmer, Emiel & Swerts, Marc (2006) Real vs. acted emotional speech.

Proceedings of the International Conference on Spoken Language Processing (Interspeech 2006), (pp. 805-808), September 2006, Pittsburgh, Pa, USA.

STEM-, SPRAAK- EN TAALPATHOLOGIE

Gepubliceerd online: 5 november 2022

University of Groningen Press
Vol. 27, 2022, Special Issue, pp. 223-249
Special Issue for Steven Gillis
https://doi.org/10.21827/32.8310/2022-SG-223

Production of pronoun gender by children acquiring Russian as a minority language: comparison with the effects of developmental language disorder

Elena Tribushinina* & Pim Mak

Utrecht Institute of Linguistics OTS, Utrecht University

Samenvatting

Dit artikel beschrijft een onderzoek naar de verwerving van het voornaamwoordelijk geslacht door Russisch-Nederlandse simultaan tweetalige kinderen, die in Nederland opgroeien. De prestaties van de tweetalige groep worden vergeleken met die van eentalige Russische kinderen met en zonder taalontwikkelingsstoornis (TOS). De groepen zijn gematcht op leeftijd. De hypothese was dat de tweetalige kinderen, die opgroeien in een context waarin het Russisch een minderheidstaal is (minder input), problemen hebben die vergelijkbaar zijn met die van kinderen met een TOS. In het laatste geval zijn de problemen een gevolg van een verminderd vermogen om taal op te nemen (minder intake). De kinderen moesten verhaaltjes vertellen aan de hand van plaatjes. De resultaten lieten zien dat beide groepen eentalige kinderen al vanaf 4 jaar nauwelijks fouten maakten in het gebruik van het geslacht van pronomina. De eentalige kinderen met een TOS waren daarbij niet te onderscheiden van hun eentalige leeftijdsgenoten zonder TOS. Dit biedt ondersteuning aan theorieën over TOS die rekening houden met de invloed van de morfologische rijkdom van een taal. In tegenstelling tot de eentalige kinderen, presteerden 4-jarige tweetalige kinderen nog op kansniveau. De prestaties bij deze groep waren beter voor oudere dan voor jongere kinderen, maar pas op 7-jarige leeftijd waren de prestaties op hetzelfde niveau als die van de eentalige kinderen. De resultaten suggereren dat in een morfologisch rijke taal een gebrek aan input meer invloed heeft op de verwerving van grammaticaal geslacht dan een gebrek aan intake. De mogelijke effecten van taalontwikkelingsstoornissen zijn na het derde levensjaar niet meer zichtbaar.

*The first author was privileged to collaborate with Steven Gillis in several projects. His cross-linguistic work on the role of morphological richness in first language acquisition has played an important role in the conceptualization of the research reported in this paper.

Correspondentieadres:
Elena Tribushinina
Utrecht University, Utrecht Institute of Linguistics OTS, Trans 10, 3512 JK, Utrecht
E-mail: e.tribushinina@uu.nl

Abstract

This paper studies the acquisition of Russian pronominal gender by Dutch-Russian simultaneous bilinguals (4;3-7;11) growing up in the Netherlands. The performance of the bilingual group is compared to that of age-matched monolinguals with and without developmental language disorder (DLD). We hypothesize that reduced exposure to Russian in the minority-language context may lead to delays in language development, comparable to problems attested in DLD (in this case due to reduced intake). The results of a narrative elicitation task demonstrate that both monolingual groups performed at ceiling from age 4 onwards. Monolingual children with DLD were as accurate at using pronominal gender as their unimpaired peers from the earliest ages studied, which supports the processing accounts of DLD taking morphological richness of the target language into account. In contrast, 4-year-old bilinguals performed around chance level. The performance of the bilingual group improved with age and reached the monolingual level only by age 7. The results suggest that reduced input has more impact on the acquisition of gender in a morphologically rich language, whereas the possible effects of DLD are no longer visible after age 3.

Introduction

Bilingual children grow up with two languages and therefore do not hear and speak each of their languages as often as their monolingual peers do. This particularly holds for the non-dominant (minority) language that bilinguals only hear and use at home and in which they receive no, or very little, schooling (Janssen, 2016; Ringblom, 2014; Tribushinina et al., 2017). Within the usage-based framework, input frequency is one of the strongest predictors of the acquisition rate (Behrens, 2009; Tomasello, 2003). Higher token frequencies leave a deeper trace in the processing system and lead to entrenchment of words and grammatical structures. Higher type frequencies bolster analogy and generalization, i.e. processes that play a paramount role in the acquisition of grammar rules (Bybee 2007; Goldberg 2006). Bilingual children usually have reduced exposure to one or both languages, which entails lower input frequencies. Not surprisingly, early bilinguals have been shown to have a slower pace in the acquisition of frequency-sensitive grammatical constructions, such as grammatical gender (Nicoladis & Marchak, 2011).

Due to the lack of reliable assessment tools for bilinguals, these children are sometimes overdiagnosed (or underdiagnosed) for developmental language disorder (DLD, formerly known as SLI) (Grimm & Schulz, 2014). Children with DLD have language deficits in the absence of any hearing, intellectual and emotional impairments or frank neurological damage (Leonard, 2014). Although there is no theoretical conformity regarding the nature of DLD (e.g. Leonard, 2014; Meir & Armon-Lotem, 2017; Rice, 2004; Ullman & Pierpont, 2005), there is a growing body of evidence suggesting that the disorder is associated with impaired procedural learning (Ullman & Pierpont, 2005) and with processing deficits, including deficits in working memory (Ellis Weismer et al., 1999) and processing speed (Windsor, 2002). In this paper, we adopt the processing approach to DLD (Ellis Weismer & Evans, 2002; Leonard, 2014), which posits that processing difficulties are cause, rather than consequence of language deficits in DLD (contra Rice, 2004). In line with this view, there is evidence that individuals with DLD need at least twice as much input to learn patterns based on statistical information in the input, compared to typically developing (TD) peers (Evans et al., 2009; Tomblin et al., 2007). Hence, in the case of monolinguals with DLD the problem is not the amount of input they receive, but the capacity to efficiently use input for acquisition (reduced intake).

Research trying to tease apart language profiles of TD bilingual children and children with DLD is clearly warranted (e.g. Armon-Lotem, 2014; Bedore & Peña, 2008; Crago & Paradis, 2003; Genesee et al., 2004; Kohnert et al., 2009; Paradis, 2010; Paradis & Crago, 2000). The present paper contributes to this literature by comparing the production of Russian pronominal gender in three groups of children: TD monolinguals, monolingual children with DLD and TD Dutch-Russian bilinguals (2L1) raised in the Netherlands. Across languages, both bilingual children (e.g. Blom et al., 2008; Blom & Vasić, 2011; Gathercole, 2002; Janssen, 2016; Kupisch et al., 2002; Nicoladis & Marchak, 2011; Rodina & Westergaard, 2017; Unsworth, 2008; 2013) and monolingual children with DLD (e.g. Anderson & Lockowitz, 2009; Bedore & Leonard, 2001; 2005; Leonard et al., 2001; Orgassa, 2009; Roulet-Amiot & Jacubowicz, 2006; Silveira, 2011) have been shown to experience difficulty in the acquisition of grammatical gender.

Studies directly comparing the performance of bilingual children and children with DLD in the gender domain report controversial results. For instance, Keij et al. (2012) have found that Dutch-speaking children with DLD lag behind TD L2 learners, which points at a greater impact of the disorder. As against this, Orgassa and Weerman (2008) have demonstrated that monolingual Dutch-speaking children with DLD perform better than TD L2 children, suggesting that reduced input in L2 children has a stronger impact than the processing deficits associated with DLD (Orgassa & Weerman, 2008). A more recent study by Marinis and colleagues (2017) compared the acquisition of gender in L2 Dutch (opaque system) and L2 Greek (transparent system). The results revealed that in Dutch both L2 children and L1 children with DLD performed worse than age-matched TD controls, whereas in Greek only TD L2 participants showed poorer performance. In contrast, L1 Greek children with DLD performed on a par with TD monolinguals. The authors conclude that problems with grammatical gender are not ubiquitous: Children with DLD acquiring morphologically rich languages have less difficulty with the acquisition of morphosyntax compared to learners of morphologically sparse languages. This is because morphological richness facilitates the acquisition of inflectional morphology by making it more salient (due to more evidence available in the input) and more transparent (due to one-to-one form-meaning mappings) (Dressler, 2005; Laaha & Gillis, 2007; Xanthos et al., 2011).

The present paper will build on this line of research and extend it to the acquisition of pronominal gender in the minority language of simultaneous bilinguals. Existing research on the effects of bilingualism and DLD tends to focus on the majority languages widely spoken in the community. In the present study, we focus on the heritage/minority/home language in which bilingual children receive much less input. We study gender production in Russian, a morphologically rich language with multiple and salient gender cues (cf. Marinis et al., 2017). Most, if not all, studies investigating the acquisition of grammatical gender by bilingual children and/or children with DLD deal exclusively with determiner and adjective agreement. An important aspect of grammatical gender that remains largely underinvestigated, also in research on typical first language acquisition, is pronominal gender, i.e., a gender agreement relation between a pronoun and its antecedent (Corbett, 1991). This study will fill this gap and focus on the acquisition of gender agreement between personal pronouns and their antecedents. By way of illustration, consider the following example from Russian:

(1) Александр подарил Марии кота. Он очень пушистый.

Alexander gave Maria cat-MASC. he very fluffy

'Alexander gave Maria a cat. It is very fluffy.'

The noun кот is masculine, hence the pronoun referring to it should also have the masculine form он 'he'. In Corbett's (1979) agreement hierarchy, pronouns involve the largest possible agreement domain because they are separated from their antecedent by large stretches of discourse. In (1) the pronoun refers to an antecedent in the preceding clause, but the referential distance between a pronoun and its antecedent can also include several clauses.

There are reasons to assume that pronominal gender might be more complex than adjective or determiner agreement because it involves not only lexical (gender assignment) and morphological (gender agreement) processes but also establishing and maintaining discourse coherence. Both bilingual children and monolingual children with DLD have been shown to exhibit problems in the domain of referential coherence (Tribushinina et al., 2017), albeit for different reasons (Mak et al., 2017). Hence, the acquisition of pronominal gender can provide useful insights into the similarities and differences between bilingual children and children with DLD. To this end, we will compare the accuracy of pronominal gender agreement in narratives produced by Dutch-Russian bilinguals and Russian monolingual children with and without DLD. A narrative task appears particularly suitable for our purposes, since it requires both maintaining complex coherence relations (i.e. the discourse dimension of pronoun use) and selecting pronouns based on the grammatical gender of their antecedent (i.e. the morphological dimension of pronoun use).

Before describing the methodology and the results of the study, we briefly review the literature on the acquisition of grammatical gender across languages and populations, with a focus on Russian.

The acquisition of grammatical gender

The acquisition of grammatical gender by typically developing (TD) monolingual children

There are large typological differences between languages as far as transparency and salience of gender systems is concerned, and these typological properties affect the rate of acquisition. Besides semantic cues (e.g. biological gender), children can rely on noun-internal (morphophonological) cues and noun-external (agreement) cues. Noun-internal cues involve predictability of gender based on the noun ending. Noun-external cues involve agreement with other elements in the sentence, such as determiners and adjectives that adjust their form according to the gender of the noun they modify. Rodina and colleagues (2020) distinguish between three types of gender systems: (i) languages with transparent gender systems (e.g. Spanish, Italian, Greek); (ii) languages with semi-transparent systems (e.g. Hebrew, Latvian, Russian); and (iii) languages with opaque systems (e.g. Dutch, Norwegian, Irish). In transparent gender systems, noun endings are unambiguously associated with specific genders (e.g. –a is strongly associated with the feminine gender in Italian). Such systems are acquired early, by age 3 or 4 (Belacchi & Cubelli, 2012; Lew-Williams & Fernald, 2007).

The acquisition of opaque gender systems is known to be a notoriously difficult and protracted process. In these languages children still make errors at age 9 (De Houwer & Gillis, 1998; Hulk & Cornips, 2006; Rodina & Westergaard, 2013; 2017; Schaerlakens & Gillis, 1987; Thomas & Gathercole, 2007). In Netherlandic Dutch, the dominant language of the bilingual participants in this study, there are very few noun-internal cues to gender assignment. Adjective agreement with the noun is informative about the noun gender only in one specific case: singular indefinite neuter nouns are modified by the bare form of an adjective (e.g. *een groot huis* ‘a big house’), whereas in all other cases attributive adjectives end in –e (e.g. *het grote huis* ‘the big house’, *de grote tafel* ‘the big table’, d*e grote tafels/huizen* ‘the big tables/houses’, *een grote tafel* ‘a big table’). The acquisition of Dutch gender mainly boils down to learning specific determiner-noun combinations (Durieux et al., 1999).

The Russian gender system can be characterized as semi-transparent: Some of the cues are highly transparent, while other cues are more ambiguous and take longer to acquire (Rodina et al., 2020). Russian has a three-way gender system of masculine, feminine and neuter. The masculine form is considered the default (Corbett & Fraser, 2000). According to Corbett (1991), about 46% of Russian nouns are masculine, 41% feminine and 13% neuter. Nouns display gender agreement with adjectives (both attributive and predicative), demonstrative pronouns, possessive pronouns and past-tense verb forms. Grammatical gender is also manifested in agreement between singular pronouns and their antecedents. Masculine nouns (e.g. стол ‘table’) require the pronoun он ‘he’, feminine nouns (e.g. картина ‘painting’) agree with the pronoun она ‘she’ and neuter nouns (e.g. окно ‘window’) with оно ‘it’ (or their oblique case forms).

Table 1: Declensional classes of Russian nouns

	I (M)			II (F)	III (F)	IV (N)
	стол 'table' (inanimate)	кот 'cat' (animate)	конь 'horse'	корова 'cow'	моль 'moth'	окно 'window'
NOM	стол-ø	скот-ø	конь-ø	коров-а	моль-ø	окн-о
GEN	стол-а	кот-а	кон-я	коров-у	мол-и	окн-а
DAT	стол-у	кот-у	кон-ю	коров-е	мол-и	окн-у
ACC	стол-ø	кот-а	кон-я	коров-у	моль-ø	окн-о
INS	стол-ом	кот-ом	кон-ём	коров-ой	моль-ю	окн-ом
LOC	стол-е	кот-е	кон-е	коров-е	мол-и	окн-е

In most cases the gender can be easily predicted from the morphophonological form of the noun (see Table 1). Masculine nouns usually end in a consonant, either hard (e.g. телефон 'telephone') or soft (e.g. медведь 'bear'). Most feminine nouns end in –a (e.g. обезьяна 'monkey'), and neuter nouns end in –o/–e (e.g. молоко 'milk', сердце 'heart'). Monolingual Russian children with typical language development acquire these major morphophonological regularities in the gender domain by age 4 (Gvozdev, 1961; Rodina, 2008).

However, some less transparent cases take longer to acquire (Rodina, 2014; Rodina & Westergaard, 2012). For example, feminine nouns of declension III end in a palatalized consonant (e.g. лошадь 'horse') and are therefore ambiguous between feminine and masculine, because there are also masculine nouns of declension I ending in a palatalized consonant (e.g. огонь 'fire'). In these cases, children have to rely on noun-external cues and noun endings in oblique cases.

Another relatively difficult aspect of the Russian gender system are nouns with a mismatch between semantic and morphological gender. Some of these nouns refer to males (semantic cue), but have a feminine form (morphological cue). For example, the noun папа 'daddy' is a feminine noun of declension II (thus having the case and number morphology of feminine nouns). However, it agrees exclusively with masculine forms of modifiers, verbs and pronouns. Similarly, female names taking the diminutive suffix –ок/ик (e.g. Светик from Света) are morphologically masculine, but should agree with feminine modifiers, verbs and pronouns.

Cases that take longest to acquire involve homophony between feminine forms and stem-stressed words. Endings of stem-stressed neuter nouns (e.g. сен-о[ə]) 'hay') are virtually indistinguishable from those of feminine nouns (e.g. машин-а[ə] 'car') because unstressed endings are always phonologically reduced in Russian. The same indistinguishability applies to stem-stressed adjectives (e.g. красн-ое[əjə] vs. красн-ая[əjə])) and stem-stressed past-tense verbs (e.g. упал-а[ə] vs. упал-о[ə])). In this case, learners have to rely on end-stressed modifiers and verbs, and also on endings in oblique cases. Six-year-olds with typical language development still make errors with stem-stressed neuter nouns, as in большая -FEM солнце-NEUT 'big sun'(Janssen, 2016; Rodina & Westergaard, 2017; Tribushinina et al., 2018).

The acquisition of pronominal gender by TD monolingual children

The vast majority of studies on the acquisition of grammatical gender deal with gender agreement marked on determiners and adjectives. Relatively little is known about the acquisition of pronominal gender. Mills (1986) reports that English-speaking preschoolers aged 3-4 correctly apply *it* to inanimate referents, but tend to over-generalize *he* to all animate referents. Similar observations were made for children acquiring Dutch (Schaerlaekens & Gillis, 1987). There is some evidence that children exposed to Netherlandic Dutch acquire basic principles of pronominal gender by age 5 (Hulk & Cornips, 2010), but gender errors persist at least until age 7 (Tribushinina & Mak, 2022).

Netherlandic Dutch is a partial pronominal language with a two-way gender system (common and neuter) in the attributive domain and a three-way gender system (masculine, feminine and neuter) in the pronominal domain. Feminine pronouns are used for human (and sometimes animal) females, masculine pronouns are used with reference to human males and animals, but also inanimate referents that are bounded and countable, and neuter pronouns are used for unbounded and uncountable inanimates (Audring, 2009a; 2009b). So an adult speaker of Netherlandic Dutch would usually refer to a fish with *hij* 'he', unless it is a cartoon fish which is conspicuously female, such as Dory in *Finding Nemo* (in this case the feminine pronoun *zij* 'she' will be used). Finally, if fish is a dish (unbounded and uncountable), the neuter pronoun *het* 'it' would be the preferred option. Dutch children as old as 7 years of age have been shown to overuse masculine pronouns, even with reference to female humans (e.g., *Hij is een prinses* 'He is a princess') and conspicuously female animals (e.g., a mother-bird) (Tribushinina & Mak, 2022).

Whereas pronominal gender in English and Netherlandic Dutch is semantic in nature (inanimate vs. animate; male vs. female), languages such as German and Belgian Dutch (also) pronominalize based on the grammatical gender of the antecedent. This form of gender agreement between the noun and the pronoun is more complex than semantic agreement and presumably takes longer to acquire. The scarce evidence available in the literature supports this assumption. For example, Mills (1986) has found that German-speaking children only learn to pronominalize correctly by age 6. Similarly, De Vogelaer (2006) reports that children speaking Belgian Dutch acquire semantic aspects of pronominal gender (based on biological sex) by age 5, but the acquisition of grammatical gender is not completed by age 7.

To the best of our knowledge, there are no studies investigating the acquisition of pronominal gender in Russian. The bilingual participants in this study acquire Russian and Netherlandic Dutch simultaneously. This language combination is interesting, since Russian has a three-way system of grammatical gender determining pronoun use, whereas in present-day Netherlandic Dutch pronoun selection is primarily based on semantic cues, as explained above.

The acquisition of grammatical gender by bilingual children

A general finding in the literature is that it takes bilingual children longer to acquire grammatical gender compared to monolingual children. Child L2 learners are almost always outperformed by their monolingual peers (e.g. Blom et al., 2008; Blom & Vasić, 2011; Brouwer et al., 2008; Hulk & Cornips, 2006; Keij et al., 2012; Orgassa & Weerman, 2008; Unsworth, 2008; Unsworth et al., 2014). For simultaneous bilinguals the findings in the literature are more mixed. Some studies report that simultaneous bilinguals make more errors with gender markers than their monolingual peers either in the majority language (Unsworth, 2008; 2013) or in the minority language (Janssen, 2016), or in both (Kupisch et al., 2002; Rodina & Westergaard, 2017). However, there are also investigations that find no differences between monolinguals and simultaneous bilinguals (Rodina & Westergaard, 2013; Unsworth et al., 2014). There is even evidence that bilinguals may outperform their monolingual peers in the acquisition of an opaque gender system if their other language has a more transparent and salient gender system (Cornips & Hulk, 2006; Tribushinina & Mak, 2022).

There are also controversial findings concerning qualitative differences between acquisition trajectories in monolingual and bilingual development. De Houwer (1990), for instance, argues that simultaneous bilinguals generally make the same errors as monolinguals. Similar findings have been reported by Blom et al. (2008) and Orgassa and Weerman (2008). In contrast, Rodina and Westergaard (2017) demonstrate that Russian-Norwegian bilinguals with very little exposure to the minority language (Russian) often make errors that monolingual Russian children do not make, such as errors with transparent nouns or across-the-board use of the default masculine forms (cf. Janssen, 2016).

Like in monolingual development, the transparency of the target gender system predicts the ease with which gender is acquired. For instance, Unsworth et al. (2014) report that Greek-English bilinguals have less difficulty acquiring gender in Greek than Dutch-English bilinguals in Dutch, since grammatical gender in Greek is much more transparent and salient than in Dutch. The amount of input in each of the child's languages is another important predictor for the ease with which gender is acquired. For example, Rodina and Westergaard (2017) report that Russian-Norwegian bilinguals (growing up in Norway) acquire the Russian gender system almost as fast as monolingual Russian children, if both their parents speak Russian at home. In contrast, children from mixed families who speak both Russian and Norwegian at home perform significantly worse with the Russian gender. Interestingly, for the acquisition of gender in the majority language (Norwegian) it apparently does not matter whether bilinguals have one or two Russian-speaking parents. In Norwegian both bilingual groups performed slightly worse than monolinguals, but did not differ from each other.

There are only two studies that looked specifically at production of gender markings by Dutch-Russian bilinguals (Janssen, 2016; Tribushinina & Mak, 2022). Janssen (2016) focused on gender production in Russian (minority language) and studied agreement between nouns and the possessive pronoun мой 'my' (in the attributive domain). Monolingual Russian-speaking children (aged 3 to 6) revealed ceiling performance across all genders. In contrast, the accuracy rates of bilingual children ranged between 55% for neuter and 67% for masculine.

Tribushinina and Mak (2022) looked at the majority language (Dutch) and compared production of pronoun gender in the narratives of Dutch-Russian bilinguals and Dutch-speaking monolinguals. The results revealed that simultaneous bilinguals outperformed Dutch monolinguals. Monolingual Dutch 7-year-olds still used masculine pronouns across the board, even with reference to female characters, such as a mother-bird. In contrast, Dutch-Russian simultaneous bilinguals appropriately used both masculine and feminine pronouns and performed like Dutch-speaking adults. This result has been taken as evidence of positive transfer from a morphologically rich minority language with a semi-transparent gender system (Russian) to a morphologically sparse majority language with an opaque gender system (Dutch). However, this study also found evidence of negative transfer from Russian: Child L2 learners of Dutch overused feminine pronouns, for instance, with reference to a fox, which is feminine in Russian and masculine in Dutch. These results show that the two gender systems do not have to be fully parallel for cross-linguistic transfer to take place: Pronoun gender is grammatical in Russian and semantic in Netherlandic Dutch. Furthermore, Russian, unlike Dutch, has no gender-marked determiners.

The present study continues this line of research and addresses production of pronoun gender in the minority language of Dutch-Russian bilinguals residing in the Netherlands. Based on the results reported in Janssen (2016), we expect that monolingual Russian-speaking children will be target-like in the production of pronoun gender by age 4, whereas Dutch-Russian bilinguals will perform significantly worse than monolinguals.

The acquisition of grammatical gender by children with DLD

As in typical language development, morphological richness of the language and transparency of its gender system have direct implications for the course of gender acquisition by individuals with DLD. Although children with DLD acquiring languages with (semi-)transparent gender systems are often outperformed by their peers with typical language development

(Anderson & Lockowitz, 2009; Bedore & Leonard, 2001; 2005; Bortolini et al., 1997; Dromi et al., 1993; Tribushinina & Dubinkina, 2012; Tribushinina et al., 2018), they still perform much better than children with DLD acquiring opaque gender systems (Keij et al., 2012; Leonard et al., 2001; Marinis et al., 2017; Orgassa & Weerman, 2008). These findings are consistent with the processing accounts of DLD that take morphological richness of the target language into account (Dromi et al., 1993; Leonard, 2000; Leonard et al., 1987). By this view, children with DLD have to devote their limited processing resources to the aspects of grammar that are most informative for interpreting sentences. For languages with sparse morphology word order is a more reliable cue, whereas for morphologically rich languages inflectional morphology is more informative. Inflections are not only more reliable but also more salient in languages with rich morphological paradigms and they offer their learners a lot of evidence of the relevant gender distinctions, thereby facilitating acquisition (Laaha & Gillis, 2007).

The only study that has addressed the acquisition of gender by Russian-speaking children with DLD is Rakhlin et al. (2014). In this study, the children (aged 7-15) were asked to assign real nouns and pseudo-nouns to either masculine or feminine gender. In addition to the morphophonological form of the noun, there were also conditions in which the children could use gender agreement with an adjective or past-tense verb as a cue to noun gender. The results revealed that children with DLD performed rather well with real nouns, but their performance with pseudo-words was below chance. These results seem to support the view that children with DLD acquire gender primarily via a lexical route, in line with the procedural deficit hypothesis (Ullman & Pierpont, 2005). By this view, individuals with DLD have impaired procedural memory (underlying the acquisition of rule-based aspects of language) but intact declarative memory (responsible for learning words and other idiosyncratic elements).

Two other studies on DLD in Russian that might be (indirectly) relevant for our present purposes are Tribushinina and Dubinkina (2012) and Tribushinina et al. (2018). These studies were not concerned with gender agreement per se, but investigated production of antonymous adjectives and adjectival degree markers by children with DLD and aged-matched TD controls. Remarkably, the children in these studies were provided with a model adjective in the correct gender form (agreeing with the target noun). They only had to provide an antonym or a degree-modified adjective in the same gender form. But even in this case, children with DLD made errors with stem-stressed neuter forms, even at the age of 10 years. These results suggest that despite the transparency of the Russian gender system, its acquisition may still be demanding for children with DLD. It remains to be seen whether this problem is specific to adjective agreement or can be generalized to other manifestations of grammatical gender. The present study will shed more light on this issue by focusing on gender agreement in the pronominal domain.

Hypotheses

This study investigates agreement between antecedent nouns and gender-marked pronouns in the Russian narratives produced by bilingual Dutch-Russian children and their monolingual peers with and without DLD (age range 4-7). Based on the literature review above, we expect that TD monolingual Russian-speaking children will be at ceiling with pronominal gender from the earliest ages studied (Gvozdev, 1961; Janssen, 2016; Rodina, 2008; Rodina & Westergaard, 2012). Both monolingual children with DLD (cf. Rakhlin et al., 2014; Tribushinina & Dubinkina, 2012; Tribushinina et al., 2018) and TD Dutch-Russian simultaneous bilinguals (cf. Janssen, 2016; Tribushinina et al., 2018) are expected to perform worse than TD Russian-speaking monolinguals. As far as the relative impact of reduced intake (DLD) vs. reduced input (2L1) is concerned, it is difficult to make specific predictions because earlier attempts to compare gender production by children with DLD and bilingual children (Keij et al., 2012; Orgassa & Weerman, 2008) only included L2 children (rather than simultaneous bilinguals), and neither of the studies looked at the acquisition of a minority language. Theoretically, three scenarios are possible.

One possibility is that there will be no difference in performance between Dutch-Russian TD bilinguals and monolinguals with DLD. For instance, a study on the production of discourse connectives by Russian-speaking children with DLD and Dutch-Russian bilinguals acquiring Russian as a heritage language (Tribushinina et al., 2017) has found that both groups made significantly more errors in connective use compared to TD monolingual peers. Furthermore, bilinguals could not be distinguished from monolinguals with DLD based on error rates and error types. A similar pattern might also emerge in the present study.

Another possibility is that 2L1 children will outperform children with DLD. This scenario is plausible because pronominal gender involves not only morphological processes (gender agreement) as such, but also discourse processes of maintaining reference across clause boundaries. There is ample evidence that that children with DLD have difficulty producing coherent discourse (e.g. Norbury et al., 2014; Tsai & Chang, 2008). If production of a coherent narrative is particularly demanding for children with DLD, their performance with pronominal gender can also be affected. An eye-tracking study testing comprehension of Russian discourse connectives (Mak et al., 2017) revealed that bilingual preschoolers were as sensitive to the semantic-pragmatic properties of these connectives as their monolingual TD peers, whereas monolingual children with DLD did not reveal any sensitivity to the discourse-organizational constraints associated with the connectives under study. These results suggest that bilingual children may have an advantage over children with DLD in the acquisition of discourse coherence phenomena, including pronominal gender.

Finally, it is possible that reduced input in simultaneous bilingualism may have greater consequences for the course of acquisition than reduced intake in DLD. In this case, monolingual children with DLD will outperform the bilingual group. Evidence in favour of this possibility comes from studies suggesting that children with DLD acquiring morphologi-

Table 2: The participants

Age group	TD monolinguals		Monolinguals with DLD		Bilinguals (2L1)	
	N	Mean age (range)	N	Mean age (range)	N	Mean age (range)
4-year-olds	24	4;3 (4;1-4;5)	20	4;6 (4;1-4;11)	25	4;4 (4;1-4;10)
5-year-olds	20	5;6 (5;4-5;10)	18	5;5 (5;2-5;11)	25	5;6 (5;0-5;11)
6-year-olds	28	6;5 (6;0-6;11)	26	6;5 (6;0-6;11)	19	6;3 (6;0-6;6)
7-year-olds	21	7;4 (7;0-7;10)	19	7;5 (7;0-7;11)	24	7;4 (7;0-7;11)

cally rich languages with transparent gender systems perform relatively well, albeit less well than TD monolinguals (Anderson & Lockowitz, 2009; Bedore & Leonard, 2001; 2005; Bortolini et al., 1997; Dromi et al., 1993). Marinis et al. (2017) found no differences between monolinguals with and without DLD in the production of gender in Greek; both monolingual groups outperformed child L2 learners. Janssen (2016) reports that Dutch-Russian preschoolers perform around chance level with gender in the attributive domain in Russian. If these accuracy rates are indicative of performance with pronominal gender, we may expect poor performance of 2L1 children below age 7 with the gender of pronouns. For this reason, we will include age groups ranging from 4 to 7 years in the present study.

Notice that the performance of the bilingual group can also be affected by cross-linguistic influence from their dominant language (Dutch). There is evidence that gender properties of one language are also activated when bilinguals use their other language (Ganushchak et al., 2011). Furthermore, young 2L1 children sometimes make gender errors that are compatible with the gender of the counterpart noun in their other language (Cantone & Müller, 2008). If the bilingual participants in our study are influenced by pronominal gender in Dutch, we may expect an overuse of masculine pronouns in their Russian narratives, since in Dutch masculine pronouns are in principle used for all non-human countable referents (Audring, 2009b).

Method

Participants

Informed consent was obtained from the parents of all participating children. Two hundred sixty-nine children participated in this study: 93 TD monolingual children, 83 monolingual children with DLD and 93 TD bilingual children. The children were divided over four age groups: 4- to 7-year-olds (see Table 2).

The bilingual Dutch-Russian participants were recruited from the Russian weekend schools in Amsterdam, Amersfoort, The Hague and Hilversum (The Netherlands). These children were born in the Netherlands and raised bilingually from birth by a Russian-speaking mother and a Dutch- or Russian-speaking father. All children attended a regular Dutch primary school on weekdays and a Russian language school/kindergarten during the weekend (half a day). In order to estimate the amount of their overall exposure to Dutch and Russian, a detailed questionnaire was administered to the parents. The analysis of the parental questionnaires reveals that all bilingual participants had less than 45% exposure to Russian (range 12–44%). The children had more exposure to Russian at the weekend (M = 12.2 hours a day) than on weekdays (M = 5.5 hours a day). On average, the children spent 2.5 weeks a year visiting relatives in Russia. The bilingual participants had no hearing problems and no history of DLD, as established by the parental questionnaire. However, eight parents reported a slow start in both Dutch and Russian. Additionally, seven parents mentioned in their comments that the Russian of their children was influenced by Dutch, as evidenced by frequent code-switches and non-Russian syntactic constructions.

The participants with DLD were recruited through specialist schools and kindergarten groups for children with DLD in a large city in West-Siberia. All children were monolingual speakers of Russian and had been independently diagnosed for DLD (in Russian – общее недоразвитие речи, уровень речевого развития III) by a multidisciplinary committee consisting of a speech pathologist, a psychiatrist, a neurologist, a paediatrician and a clinical psychologist. For privacy reasons we were not granted access to the diagnostic results. Therefore, the teachers were asked to select participants based on the following set of criteria: medium- to high performance on a series of non-verbal IQ tests conducted by a school/kindergarten psychologist; no evidence of neurological impairment; no severe visual or auditory problems (based on the yearly medical checks at kindergarten/school); absence of any other known disorder, such as autism; no severe phonological disorder.

The monolingual TD Russian group was recruited from a number of kindergartens and primary schools in the same city as the DLD group. The teachers were asked to select the children based on the following criteria: normal IQ (within one standard deviation of the mean on IQ tests conducted for school enrolment), average academic performance, normal motor, social-emotional and cognitive development, as well as age-appropriate language skills.

Materials and procedure

Two picture stories were used to elicit children's narratives - the Fox Story (Gülzow & Gagarina, 2007) and the Cat Story (Hickmann, 2003). The stories consisted of six pictures each. The pictures were black-and-white drawings, 12x12 cm (Fox Story) and 10x13 cm (Cat Story) in size. In the Cat Story, there are three main protagonists – a mother-bird, a cat and a dog. The mother-bird flies away from the nest in order to search for food for her nestlings. The cat notices this and climbs the tree to catch the baby birds. Then a dog comes into the picture and just before the cat reaches the nest, the dog grabs his tail and pulls the cat from the tree. The mother-bird comes back with a worm and the dog chases the cat away from the tree with the nest. In the Fox Story there are also three main characters – a bird, a fox and a fish (fishbone). The bird finds the fish and proudly sits on a tree. A hungry fox appears and wants to have the fish. The fox reaches for the fish and the bird drops it. The fox catches the fish and runs away, but the bird follows the fox and takes the fish back.

As narrative collection was part of a larger study (see the Discourse BiSLI corpus in the CHILDES archive), the bilingual participants produced one narrative in Russian and one in Dutch (either Cat or Fox). Only Russian narratives are relevant to the present study. The monolingual participants were randomly assigned to one of the narratives (Cat or Fox) so that their corpus size would be comparable to that of the bilingual participants.

The children were interviewed individually in a quiet room. The monolingual children were tested in their regular school/kindergarten. Most bilingual participants were tested at their weekend (Russian) school, and some were tested at home. After a short warming-up talk, the investigator asked the child to tell the story in pictures. All six pictures were then placed on the table and the child was asked to look through them and confirm that they have understood the story. After that the experimenter took all the pictures away and started placing them on the table one by one. When the child finished describing picture 1, the investigator placed picture 2 next to picture 1 so that the child could see two pictures at the same time. When the child was finished with picture 2, the investigator placed it on top of picture 1 and put picture 3 next to picture 2 on the table, etc. The narratives were audio-recorded and later transcribed in a CHAT format (MacWhinney, 2000).

Coding

First, we selected all sentences containing personal pronouns, either in the nominative case (e.g. он 'he') or in one of the oblique cases (e.g. ей-DAT 'her' его-ACC 'him'). Only singular pronouns were selected because gender is not marked in plural. Demonstrative and relative pronouns were not included in this study because they have adjectival forms of gender agreement and this investigation only focussed on pronominalization. Repetitions and reformulations were left out of the analyses. Finally, only cases with clear antecedents were included in the study; we excluded all ambiguous pronouns (i.e. pronouns that could not be unambiguously related to one specific referent). For example, there are two possible antecedents of она in (2), and it is not clear from the context whether the child refers to the bird or to the cat.

(2) А потом кис&, а потом птичка говорит... А ! Заметила кошечка. А потом она, а потом она куда-то ушла и снова увидела птичек. (r_dld4_119_cat)

'And then the cat, and then the bird says.... Ah! Noticed the cat (or: the cat noticed). And then she, and she left somewhere and saw the little birds again.'

All remaining (relevant) personal pronouns (N=886) were coded for grammatical gender. There was only one instance of a neuter pronoun in the whole corpus, hence only masculine and feminine forms were included in further analyses. Table 3 summarizes the number of coded pronoun instances by group and age.

Table 3: Number of cases of relevant personal pronouns, split by group and age

	4 years	5 years	6 years	7 years
Monolingual (TD)	55	84	213	93
Monolingual (DLD)	29	28	62	51
Bilingual	60	72	75	47

Additionally, each pronoun was coded as either correct gender or incorrect gender. The target gender was determined for each individual case separately, depending on the antecedent of the personal pronoun. For instance, if the child referred to the bird in the Fox Story as птица-FEM 'bird' or ворона-FEM 'crow', the feminine pronouns would be coded as correct. However, if the same protagonist was called ворон-MASC 'raven' or орёл-MASC 'eagle', the masculine pronouns would be considered the target form.

Data analysis

The dependent variable in the analysis was whether or not a child used the correct gender of the pronoun when referring to its antecedent. We analysed the data with a logistic regression using the *glmer* function in the lme4 package in R (Bates et al., 2015). Participant and Story were included as random factors. Group (TD monolingual children vs. bilingual children vs. monolingual children with DLD), Age, and the interaction of Group and Age were included in the fixed part of the model. Age was treated as a continuous variable. The TD monolingual children were taken as the reference level, as we expected this group to show the best performance and wanted to compare the performance of the other groups to this reference group. This analysis tests whether the probability that a child uses the correct pronoun is lower for the bilingual children and the monolingual children with DLD than for the TD monolingual children. Also, the interaction tests whether the difference, if any, decreases with age.

Results

The results are presented in Figure 1.

The results of the logistic regression are presented in Table 4. There was no difference between the children with DLD and the TD monolingual children. There was no effect of Age for the monolingual groups: Both groups were at ceiling with the production of pronominal gender across all ages studied. The bilingual children performed significantly worse than the monolingual children. At age 4, when the monolingual groups were already target-like in the production of pronominal gender, the bilingual children performed just above chance. In contrast, the performance of the bilingual children increased with age. As is ev-

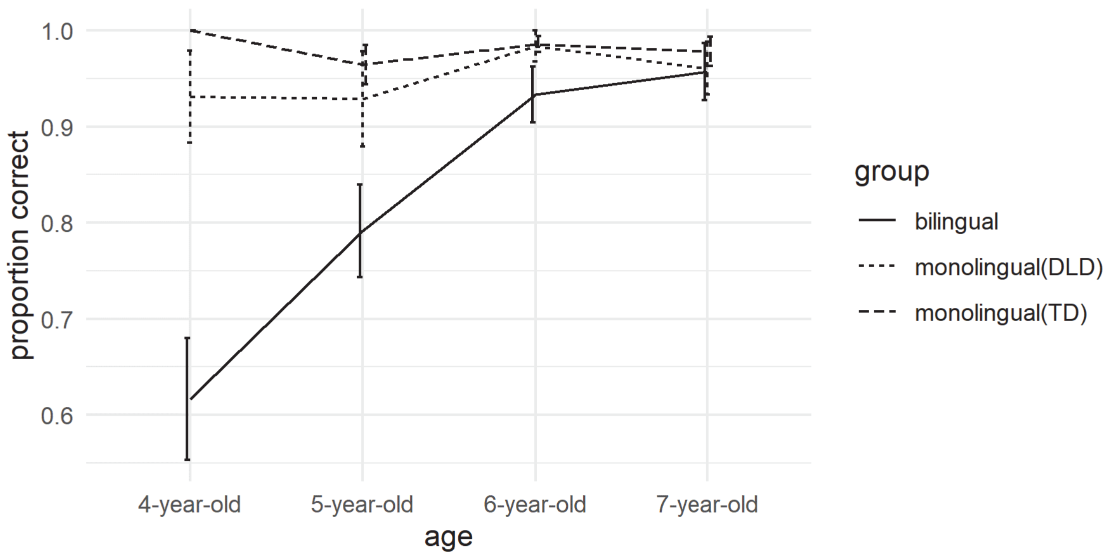


Figure 1: Proportion correct pronouns, by group and age

Table 4: Fixed effects in the logistic regression

	B	Std. Error	*z*-value	*p*
Intercept	7.33	3.84	1.91	.06
Age (L1)	-0.23	0.63	-0.36	.71
L1-DLD vs. L1-TD	-3.80	4.94	-0.77	.44
2L1 vs. L1-TD	-11.47	4.51	-2.54	.01
Age (L1-DLD vs. L1-TD)	0.50	0.85	0.59	.56
Age (2L1 vs. L1-TD)	1.57	0.78	2.02	.04

ident from Table 3, the youngest (4-year-old) bilinguals performed significantly worse than TD monolinguals. Posthoc pairwise comparisons in the other age groups revealed that at age 5 ($p = .02$) and 6 ($p = .03$) the performance of the bilingual group still differed from that of TD monolinguals. Only at age 7 there was no difference between the bilinguals and the TD monolinguals ($p = .78$).

A qualitative error analysis revealed that bilinguals made only one type of error – using the masculine pronouns instead of the feminine ones, as in examples (3) and (4). Errors in the opposite direction (feminine instead of masculine) were not attested, whereas feminine pronouns were more frequent in the narratives (69% feminine pronouns, 31% masculine pronouns).

(3) Лиса. Он хочет съесть. (br4_098_fox)
fox-FEM. he-MASC&ERR wants eat
'Fox. He wants to eat.'

The masculine pronoun он in (3) is used to refer back to the feminine noun лиса 'fox'. Similarly, the masculine accusative form его 'him' is used in (4) with reference to the cat for which the child uses the feminine noun кошка throughout the narrative.

Analysis of individual performance reveals that some of the bilinguals used both correct feminine and incorrect masculine pronouns. In total, there were 12 children in the bilingual group who showed such inconsistent performance (3 four-year-olds, 5 five-year-olds, 3 six-year-olds and 1 seven-year-old). Nine of these children used feminine and masculine pronouns interchangeably with reference to the same protagonists. For example, the pronouns он and она in (5a) and (5b) respectively were used with reference to the same character – bird. The feminine noun is appropriate in this case because the noun птица 'bird' used by this child is feminine.

(5 a.) И потом он, что-то надо делать. (br6_021_cat)
and then he-MASC&ERR something needs do
'And then he needs to do something.'

(5 b.) Она кушать взяла им. (br6_021_cat)
she-FEM eat took them
'She took some food for them.'

Interestingly, in most cases where the children switch between two gender forms for the same referent, they adequately adjust the gender form of the verb or attributive phrase; see for example (6):

(6 a.) Сначала птица нашла рыбу такую. (br4_194_fox)
first bird-FEM found- FEM fish- FEM such- FEM
'First the bird found such a fish.'
(6 b.) А потом он уронил.
and then he dropped-MASC
'And then he dropped (it).'
(6 c.) Лиса тогда его взяла и убежала.
fox- FEM then him took- FEM and ran.away- FEM
'The fox then grabbed it and ran away.'

The sentences in (6) were produced within the same narrative. Notice that in (6a) there is correct agreement between the feminine noun птица 'bird' and the verb, whereas the masculine pronoun он is erroneously used to refer to the bird in (6b). At the same time, the masculine pronoun он in (6b) correctly agrees with the verb уронил 'dropped-MASC'.

In (6a) the feminine noun рыба 'fish' correctly agrees with the demonstrative determiner, but in (6c) the fish is referred to by means of the masculine pronoun его 'him'. A similar pattern was observed in other narratives with inconsistent use of masculine and feminine forms with reference to the same protagonists.

Discussion and conclusion

This study compared production of gender agreement in the pronominal domain across three groups of children: simultaneous Dutch-Russian bilinguals with limited exposure to Russian and Russian-speaking monolinguals with and without DLD. We predicted that both 2L1 children acquiring language in the minority situation (the lower-input group) and monolingual Russian children with DLD (the lower-intake group) would have more difficulty acquiring gender agreement in the pronominal domain compared to unimpaired monolingual children. Concerning the relative impact of simultaneous bilingualism (in the minority language) and processing deficits in DLD, three possibilities were considered: no difference, greater impact of bilingualism and greater impact of the disorder.

The first prediction was only partly borne out by the data in this study. Only bilingual children lagged behind their monolingual TD peers in the acquisition of pronominal gender. Monolingual children with DLD had a ceiling performance and were indistinguishable from their unimpaired peers from the earliest ages studied. These findings are consistent with prior research demonstrating that the acquisition of gender agreement in the adjectival and verbal domain by TD Russian-speaking children is by and large completed by age 4 (Gvozdev, 1961; Janssen, 2016; Rodina, 2008). This study extends these findings to pronominal gender and to language development of children with DLD. One might assume that pronominal gender is more demanding than gender agreement with attributes and verbs because pronominal agreement extends beyond clause boundaries and requires maintaining referential coherence. However, we have seen no evidence of this in our research. Both monolingual groups showed target-like performance from age 4 onwards. Notice, however, that based on these data we cannot rule out the possibility that children with DLD have a delay in the acquisition of (pronominal) gender before age 4. It appears difficult to test this possibility because the youngest age at which children are diagnosed with DLD is 4 years.

Our results are compatible with the processing accounts of DLD (Dromi et al., 1993; Leonard, 2000; Leonard et al., 1987), which claim that children with DLD have limited processing resources and devote them to the aspects of grammar that are most informative in a given language. Inflectional morphology is a reliable cue in morphologically rich languages, such as Russian. The current results have revealed that 4-year-old Russian-speaking children with DLD are target-like in the production of pronominal gender, whereas German-speaking children with typical language development only reach this milestone by age 6 (Mills, 1986) and children acquiring Belgian Dutch still struggle with grammatical pronoun gender at age 7 (De Vogelaer, 2006). The current study adds to the body of research showing that children with DLD acquiring morphologically rich languages have an advantage over their peers speaking languages with sparse morphology (Dromi et al,. 1993; Leonard, 2000; Leonard et al., 1987; Marinis et al., 2017), even though in morphologically rich languages, such as Turkish, children with DLD may still perform worse than their TD counterparts (Güven & Leonard, 2020, 2021).

Concerning the question whether language learning in a minority context or language deficits associated with DLD would have more impact on the development of pronominal gender in Russian, the results unambiguously show that the former is the case. The performance of Russian-speaking monolinguals with DLD is not different from that of TD monolinguals. In contrast, bilingual children having limited exposure to Russian in the dominant Dutch environment significantly lag behind their monolingual peers in the acquisition of pronominal gender. At age 4, the performance of monolinguals is at ceiling, whereas bilinguals of this age still perform around chance. With age, their production of pronominal gender becomes more target-like, and at age 7 bilinguals reach the monolingual standard in the production of pronominal gender. These findings are consistent with the results reported by Marinis et al. (2017): In their study Greek-speaking monolinguals with DLD performed like TD monolinguals, whereas child L2 learners made significantly more errors than both monolingual groups.

A clinical implication of these findings is that pronoun gender, unlike gender agreement between adjectives and nouns (Tribushinina & Dubinkina, 2012; Tribushinina et al., 2018), is not vulnerable in Russian-speaking children with DLD. So poor performance on pronominal gender by pre-school children is likely to be indicative of negative effects of reduced input in the target language rather than language disorder. Given the vulnerability of pronoun gender in the minority language context, Russian weekend schools may need to include more practice with anaphoric pronouns and noun gender in their curricula. This being said, the transparency of the Russian gender system makes it possible for bilinguals with limited input in Russian to acquire the system completely by age 7, whereas bilinguals acquiring opaque gender systems may fossilize in their non-target-like performance, even in the majority language (Hulk & Cornips, 2006; Unsworth, 2008), but even more so in the minority language (Nic Fhlannchadha & Hickey, 2017; Thomas & Gathercole, 2007).

The question arises whether the poor performance of the bilingual children below age 7 is only due to limited exposure to the minority language or also due to cross-linguistic influence from their dominant language (Dutch). We cannot tease these two possibilities apart based on the production data in this study. On the one hand, given the earlier findings that Dutch-Russian bilinguals aged 3-6 perform slightly above chance in the production of gender markings on possessive pronouns (Janssen, 2016) we may assume that the errors attested in this study are, at least, partly due to a delayed acquisition of grammatical gender in the minority-language context. This explanation is in line with the earlier findings that Russian monolingual children with typical language development (Ceitlin, 2000) and with DLD (Rakhlin et al., 2014) also over-use the default masculine form.

On the other hand, it is also possible that production of pronominal gender by Dutch-Russian bilinguals might be aggravated by transfer errors. Importantly, all the errors in the bilingual group were overgeneralizations of the masculine pronouns. The bilinguals overused masculine forms for feminine antecedents, but never the other way around. It is plausible to assume that this performance might be a manifestation of cross-linguistic influence from Dutch. All the protagonists in the narratives under study (except the mother-bird) would be referred to by means of masculine pronouns in Dutch, since all bounded countable entities including animals fall within the scope of the pronoun *hij* 'he' in Netherlandic Dutch.

Based on the production of pronouns, we cannot say for sure whether the children fall back on the masculine forms due to a delayed acquisition (i.e. incomplete grammar) or over-rely on masculine pronouns due to performance problems such as inhibiting the dominant Dutch system (cf. White, 2011). It is possible that both factors converge and thereby create a combined difficulty effect. Online processing experiments might be helpful for teasing these two possibilities apart. Several investigations of bilingual acquisition of Dutch grammatical gender have shown that bilingual children who are still at chance in the production of the Dutch neuter determiner reveal sensitivity to ungrammaticality in online tasks, as evidenced in longer reaction times after ungrammatical determiner-noun combinations (Blom & Vasić, 2011; Brouwer et al., 2008) and appropriate forced-choice (offline) grammaticality judgments (Unsworth, 2013). These studies conclude that frequent errors in the production of grammatical gender may reflect "a production-specific performance problem rather than a failure to acquire those grammatical features and rules and/or to specify certain nouns with the target gender features" (Unsworth, 2013, p. 105). These findings make it plausible to assume that the problems in the production of pronominal gender by bilinguals under age 7 are at least partly due to difficulty with inhibiting the dominant Dutch system in a language production task. If this is the case, online experiments with children aged 4-5 (i.e. the age when they still perform around chance level in gender production) may reveal early sensitivity to gender cues in language processing. Such patterns would be compatible with theories of bilingual development assuming that comprehension-production asymmetries are common in bilinguals and that transfer errors are an epiphenomenon of speech production rather than indications of deviant representation (Nicoladis, 2006).

This study has a number of limitations. First, the narratives contained only a limited number of referents and the participants were free to choose any noun to refer to the protagonists. We cannot exclude the possibility that the performance of children with DLD could have been less target-like if a wider range of nouns, including more ambiguous cases, such as stem-stressed neuter nouns, were included (cf. Tribushinina & Dubinkina, 2012). Future studies should test this possibility using elicited production experiments including different noun classes. Second, we were not given access to the diagnostic results in the DLD group, so we did not have much insight into the language profiles of the participants with DLD, which is important given the heterogeneity of this population. It would also be useful to have more information on the cognitive skills (e.g. working memory, procedural memory) of participants with (and without) DLD. In this paper, we assumed that DLD involves reduced intake, but we have not directly tested this possibility. Future work in this area should empirically test the contribution of input quantity and processing skills on the acquisition of gender distinctions. Finally, as discussed above, based on production data alone, we cannot determine the source(s) of errors that bilingual children make. Online experiments (e.g. in the Visual World Paradigm) will be crucial to resolving these issues in future research.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable constructive comments. We are also very grateful to all children, parents and teachers who have made this investigation possible. Elena Dubinkina-Elgart kindly helped with data collection in Russia. This research was supported by a Marie Curie International Research Staff Exchange Scheme Fellowship within the 7th European Community Framework Programme (grant number 269173).

References

Anderson, R. T., & Lockowitz, A. (2009). How do children ascribe gender to nouns? A study of Spanish-speaking children with ad without specific language impairment. *Clinical Linguistics and Phonetics, 23*(7), 489-506. https://doi.org/10.1080/02699200902844818

Armon-Lotem, S. (2014). Between L2 and SLI: Inflections and prepositions in the Hebrew of bilingual children with TLD and monolingual children with SLI. *Journal of Child Language, 41*, 3-33. https://doi.org/10.1017/s0305000912000487

Audring, J. (2009a). Gender assignment and gender agreement: Evidence from pronominal gender languages. *Morphology, 18*(2), 93-116. https://doi.org/10.1007/s11525-009-9124-y

Audring, J. (2009b). *Reinventing pronoun gender*. LOT Dissertation Series.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bedore, L. M., & Leonard, L. B. (2001). Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*(4), 905–924. https://doi.org/10.1044/1092-4388(2001/072)

Bedore, L. M., & Leonard, L. B. (2005). Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. *Applied Psycholinguistics, 26*(2), 195-225. https://doi.org/10.1017/s0142716405050149

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11*(1), 1-29.

https://doi.org/10.2167/beb392.0
Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics, 47*(2): 383-411.
Belacchi, C., & Cubelli, R. (2012). Implicit knowledge of grammatical gender in preschool children. *Journal of Psycholinguistic Research, 41*(4), 295-310. https://doi.org/10.1007/s10936-011-9194-y
Blom, E., Polišenská, D., & Weerman, F. (2008). Articles, adjectives and age of onset: The acquisition of Dutch grammatical gender. *Second Language Research, 24*(3), 297-331. https://doi.org/10.1177/0267658308090183
Blom, E., & Vasić, N. (2011). The production and processing of determiner–noun agreement in child L2 Dutch. *Linguistic Approaches to Bilingualism, 1*(3), 265-290. https://doi.org/10.1075/lab.1.3.03blo
Bortolini, U., Caselli, M. C., & Leonard, L. B. (1997). Grammatical deficits in Italian-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 40*(4), 809-820. https://doi.org/10.1044/jslhr.4004.809
Brouwer, S., Cornips, L., & Hulk, A. (2008). Misrepresentation of Dutch neuter gender in older bilingual children? In E. Gavrusheva & B. Haznedar (Eds.), *Trends in child second language acquisition* (pp. 83-96). John Benjamins. https://doi.org/10.1075/lald.46.06bro
Bybee, J. (2007). *Frequency of use and the organization of language.* Oxford University Press.
Cantone, K. F., & Müller, N. (2008). Un nase or una nase? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Lingua, 118*(6), 810-826. https://doi.org/10.1016/j.lingua.2007.05.007
Ceitlin, S. N. (2000). *Jazyk i rebënok. Lingvistika detskoj reči.* Moskva: Vlados.
Corbett, G. G. (1979). The agreement hierarchy. *Journal of Linguistics, 15*(2), 203-224.
Corbett, G. G. (1991). *Gender.* Cambridge University Press. https://doi.org/10.1017/s0022226700015449
Corbett, G. G., & Fraser, N. M. (2000). Default genders. In B. Unterbeck (Ed.), *Gender in grammar and cognition* (pp. 55-91). Mouton.
Cornips, L. M. E. A., & Hulk, A. (2006). External and internal factors in bilingual and bidialectal language development: Grammatical gender of the Dutch definite determiner. In C. Lefebvre, L. White, & C. Jourdan (Eds.), *L2 acquisition and creole genesis* (pp. 355-378). John Benjamins. https://doi.org/10.1075/lald.42.21cor
Crago, M., & Paradis, J. (2003). Two of a kind? Commonalities and variation in languages and language learners. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Towards a definition of specific language impairment* (pp. 97–110). Lawrence Erlbaum Associates.
De Houwer, A. (1990). *The acquisition of two languages from birth: A case study.* Cambridge University Press.
De Houwer, A., & Gillis, S. (1998). *The acquisition of Dutch.* John Benjamins.
De Vogelaer, G. (2006). Pronominaal genus bij 'Zuid-Nederlandse' taalverwervers: Van grammaticaal naar semantisch systeem. In M. Hüning, U. Vogl, T. van der Wouden, & A. Verhagen (Eds.), *Nederlands tussen Duits en Engels* (pp. 89-102). Stichting Neer-

landistiek Leiden.

Dressler, W. U. (2005). Morphological typology and first language acquisition: Some mutual challenges. In G. Booij, E. Guevara, A. Ralli, S. Sgroi, & S. Scalise (Eds.), *Morphology and linguistic typology. On-line proceedings of the Fourth Mediterranean Morphology Meeting (MMM4) Catania 21-23.9.2003* (pp. 7–20). University of Bologna.

Dromi, E., Leonard, L. B., & Shteiman, M. (1993). The grammatical morphology of Hebrew-speaking children with specific language impairment: Some competing hypotheses. *Journal of Speech, Language, and Hearing Research, 36*(4), 760-771. https://doi.org/10.1044/jshr.3604.760

Durieux, G., Daelemans, W., & Gillis, S. (1999). On the arbitrariness of lexical categories. In van Eynde, F. (Ed.), *Computational Linguistics in the Netherlands 1998* (pp. 19-36). Rodopi.

Ellis Weismer, S., & Evans, J. L. (2002). The role of processing limitations in early identification of specific language impairment. *Topics in Language Disorders, 22*(3), 15-29.

Ellis Weismer, S., Evans, J., & Hesketh, L. (1999). An examination of verbal working memory capacity in children with SLI. *Journal of Speech, Language, and Hearing Research, 42*, 1249–1260.

Evans, J., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 321-335.

Ganushchak, L. Y., Verdonschot, R. G., & Schiller, N. O. (2011). When leaf becomes neuter: Event-related potential evidence for grammatical gender transfer in bilingualism. *Neuroreport, 22*(3), 106-110. https://doi.org/10.1097/wnr.0b013e3283427359

Gathercole, V. C. M. (2002). Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In D. Kimbrough Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 207-219). Multilingual Matters.

Genesee, F., Paradis, J., & Crago, M. (2004). *Dual language development and disorders: A handbook of bilingualism and second language learning.* Brookes.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language.* Oxford University Press.

Grimm, A., & Schulz, P. (2014). Specific language impairment and early second language acquisition: The risk of over-and underdiagnosis. *Child Indicators Research, 7*(4), 821-841. https://doi.org/10.1007/s12187-013-9230-6

Gülzow, I., & Gagarina, N. (2007). Noun phrases, pronouns and anaphoric reference in young children narratives. In D. Bittner & N. Gagarina (Eds.), *Intersentential pronominal reference in child and adult languag*e (pp. 203-223). ZAS Papers in Linguistics.

Güven, S., & Leonard, L. B. (2020). Production of noun suffixes by Turkish–speaking children with developmental language disorder and their typically developing peers. *International Journal of Language and Communication Disorders, 55*(3), 387-400. https://doi.org/10.1111/1460-6984.12525.

Güven, S. & Leonard, L. B. (2021) Verb morphology in Turkish-speaking children with and without DLD: the role of morphophonology, *Clinical Linguistics & Phonetics,* DOI: 10.1080/02699206.2021.2012836.

Gvozdev, A. N. (1961). *Formirovanie u rebeka grammatičeskogo stroja russkogo jazyka.* APN RSFSR.
Hickmann, M. (2003). *Children's discourse. Person, space, and time across languages.* Cambridge University Press.
Hulk, A., & Cornips, L. (2006). Neuter gender and interface vulnerability in child L2/2L1 Dutch. In S. Unsworth, T. Parodi, A. Sorace, & M. Young-Scholten (Eds.), *Paths of development in L1 and L2 acquisition (Language Acquisition and Language Disorders 39)* (pp. 107-134). John Benjamins.
Hulk, A., & Cornips, L. (2010). The role of gender and count features in the acquisition of 'het' as a pronoun: similarities and differences with its acquisition as a determiner. In J. Costa, A. Castro, M. Lobo, & F. Pratas (Eds.), *Language acquisition and development: Proceedings of GALA 2009* (pp. 229–239). Cambridge Scholars.
Janssen, B. (2016). *The acquisition of gender and case in Polish and Russian.* Pegagus.
Keij, B., Cornips, L. M., Van Hout, R., Hulk, A., & Van Emmerik, J. (2012). Knowing versus producing: The acquisition of grammatical gender and the definite determiner in Dutch by L1-TD, L1-SLI, and eL2 children. *Linguistic Approaches to Bilingualism, 2*(4), 379-403. https://doi.org/10.1075/lab.2.4.02kei
Kohnert, K., Windsor, J., & Ebert, K. D. (2009). Primary or "specific" language impairment and children learning a second language. *Brain and Language, 109*(2-3), 101-111. https://doi.org/10.1016/j.bandl.2008.01.009
Kupisch, T., Müller, N., & Cantone, K. F. (2002). Gender in monolingual and bilingual first language acquisition: Comparing Italian and French. *Lingue e Linguaggio, 1*(1), 107-150.
Laaha, S., & Gillis, S. (Eds.) (2007). *Typological perspectives on the acquisition of noun and verb morphology*. Antwerp University.
Leonard, L. B. (2000). Specific language impairment across languages. In D. V. M. Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome* (pp. 115-129). Psychology Press.
Leonard, L. B. (2014). *Children with specific language impairment.* MIT Press.
Leonard, L. B., Sabbadini, L., Leonard, J. S., & Volterra, V. (1987). Specific language impairment in children: A cross-linguistic study. *Brain and Language, 32*(2), 233-252. https://doi.org/10.1016/0093-934x(87)90126-x
Leonard, L. B., Salameh, E. K., & Hansson, K. (2001). Noun phrase morphology in Swedish-speaking children with specific language impairment. *Applied Psycholinguistics, 22*(4), 619-639. https://doi.org/10.1017/s0142716401004076
Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 18*(3), 193-198.
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Lawrence Erlbaum Associates.
Mak, W. M., Tribushinina, E., Lomako, J., Gagarina, N., Abrosova, E., & Sanders, T. (2017). Connective processing by bilingual children and monolinguals with specific language impairment: Distinct profiles. *Journal of Child Language, 44*(2), 329-345.

https://doi.org/10.1017/s0305000915000860
Marinis, T., Chondrogianni, V., Vasic, N., Weerman, F., & Blom, E. (2017). The impact of transparency and morpho-phonological cues in the acquisition of grammatical gender in sequential bilingual children and children with Specific Language Impairment. In E. Blom, L. Cornips & J. Schaeffer (Eds.), *Cross-linguistic influence in bilingualism: In honor of Aafke Hulk* (pp. 153-180). John Benjamins.
Meir, N. & Armon-Lotem, S. (2017). Delay or deviance: Old question – new evidence from bilingual children with specific language impairment (SLI). In M. LaMendola & J. Scott (Eds.), *Proceedings of the 41st Annual Boston University Conference on Language Development* (pp. 495-508). Cascadilla Press.
Mills, A. E. (1986). *The acquisition of gender a study of English and German.* Springer-Verlag.
Nic Fhlannchadha, S., & Hickey, T. M. (2017). Acquiring an opaque gender system in Irish, an endangered indigenous language. *First Language, 37*(5), 475-499. https://doi.org/10.1177/0142723717702942.
Nicoladis, E. (2006). Cross-linguistic transfer in adjective-noun strings by preschool bilingual children. *Bilingualism: Language and Cognition, 9*(1), 15-32.
Nicoladis, E., & Marchak, K. (2011). Le carte blanc or la carte blanche? Bilingual children's acquisition of French adjective agreement. *Language Learning, 61*(3), 734-758. https://doi.org/10.1111/j.1467-9922.2011.00660.x
Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language, 41*(3), 485-510. https://doi.org/10.1017/s030500091300007x
Orgassa, A. (2009). *Specific language impairment in a bilingual context: The acquisition of Dutch inflection by Turkish-Dutch learners.* LOT.
Orgassa, A., & Weerman, F. (2008). Dutch gender in specific language impairment and second language acquisition. *Second Language Research, 24*(3), 333-364. https://doi.org/10.1177/0267658308090184
Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Applied Psycholinguistics, 31*(2), 227-252. https://doi.org/10.1017/s0142716409990373
Paradis, J., & Crago, M. (2000). Tense and temporality: A comparison between children learning a second language and children with SLI. *Journal of Speech, Language, and Hearing Research, 43*(4), 834-847. https://doi.org/10.1044/jslhr.4304.834
Rakhlin, N., Kornilov, S. A., & Grigorenko, E. L. (2014). Gender and agreement processing in children with Developmental Language Disorder. *Journal of Child Language, 41*(2), 241-274. https://doi.org/10.1017/s030500091200058x
Rice, M. L. (2004). Growth models of developmental language disorders. In M.L. Rice & S. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 207-240). Lawrence Erlbaum.
Ringblom, N. (2014). *The acquisition of Russian in a language contact situation: A case study of a bilingual child in Sweden.* [Doctoral dissertation, Stockholm University].
Rodina, Y. (2008). *Semantics and morphology: The acquisition of grammatical gender in*

Russian. [Doctoral dissertation, University of Tromsø].

Rodina, Y. (2014). Variation in the input: Child and caregiver in the acquisition of grammatical gender in Russian. *Language Sciences, 43*, 116-132. https://doi.org/10.1016/j.langsci.2013.10.007

Rodina Y., Kupisch, T., Meir, N., Mitrofanova, N., Urek, O., & Westergaard, M. (2020). Internal and external factors in heritage language acquisition: Evidence from heritage Russian in Israel, Germany, Norway, Latvia and the United Kingdom. *Frontiers in Education, 5*, 20. https://doi.org/10.3389/feduc.2020.00020

Rodina, Y., & Westergaard, M. (2012). A cue-based approach to the acquisition of grammatical gender in Russian. *Journal of Child Language, 39*(5), 1077-1106. https://doi.org/10.1017/s0305000911000419

Rodina, Y., & Westergaard, M. (2013). The acquisition of gender and declension class in a non-transparent system: Monolinguals and bilinguals. *Studia Linguistica, 67*(1), 47-67. https://doi.org/10.1111/stul.12012

Rodina, Y., & Westergaard, M. (2017). Grammatical gender in bilingual Norwegian–Russian acquisition: The role of input and transparency. *Bilingualism: Language and Cognition, 20*(1), 197-214. https://doi.org/10.1017/s1366728915000668

Roulet-Amiot, L., & Jakubowicz, C. (2006). Production and perception of gender agreement in French SLI. *Advances in Speech Language Pathology, 8*(4), 335-346. https://doi.org/10.1080/14417040601009420

Schaerlaekens, A.M., & Gillis, S. (1987). *De taalverwerving van het kind.* Wolters-Noordhoff.

Silveira, M. (2011). *Specific Language Impairment (SLI) revisited: Evidence from a psycholinguistic investigation of grammatical gender abilities in Brazilian Portuguese - speaking children.* [Doctoral dissertation, University College London].

Thomas, E. M., & Gathercole, V. C. M. (2007). Children's productive command of grammatical gender and mutation in Welsh: An alternative to rule-based learning. *First Language, 27*(3), 251-278. https://doi.org/10.1177/0142723707077056

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press.

Tomblin, B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in children with and without specific language impairment. *Journal of Child Language Learning and Development*, 3, 269-293.

Tribushinina, E. and Dubinkina, E. (2012). Adjective production by Russian-speaking children with specific language impairment. *Clinical Linguistics and Phonetics, 26*(6), 554-571.

Tribushinina, E., & Mak, P. (in press, 2022). Linguistic advantages of bilingualism: The acquisition of Dutch pronominal gender. In G. Cascio, O. Secardin, R. Supheert, & J. ten Thije (Eds.), *Crossing borders: Challenging linguistic and cultural diversity* (pp. 60-76). Brill.

Tribushinina, E., Mak, W. M., Andreiushina, E., Dubinkina, E., & Sanders, T. (2017). Connective use in the narratives of bilingual children and monolingual children with SLI. *Bilingualism: Language and Cognition, 20*(1), 98-113. https://doi.org/10.1017/s1366728915000577

Tribushinina, E., Mak, M., Dubinkina, E., & Mak, W.M. (2018). Adjective production by Russian-speaking children with developmental language disorder and Dutch–Russian simultaneous bilinguals: Disentangling the profiles. *Applied Psycholinguistics*, 1-32. doi:10.1017/S0142716418000115

Tsai, W., & Chang, C. J. (2008). “But I first... and then he kept picking”: Narrative skill in Mandarin-speaking children with language impairment. *Narrative Inquiry, 18*(2), 349-377. https://doi.org/10.1075/ni.18.2.09tsa

Ullman, M., & Pierpont, R. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex, 41*, 399–433.

Unsworth, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second Language Research, 24*(3), 365-395. https://doi.org/10.1177/0267658308090185

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition, 16*(1), 86-110. https://doi.org/10.1017/s1366728912000284

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics, 35*(4), 765-805. https://doi.org/10.1017/s0142716412000574

White, L. (2011). Second language acquisition at the interfaces. *Lingua, 121*(4), 577-590.

Windsor, J. (2002). Contrasting general and process-specific slowing in language impairment. *Topics in Language Disorders, 22*(3), 49-61.

Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., Gagarina, N., Hrzica, G., Ketrez, F.N., Kilani-Schoch, M., Korecky-Kröll, K., Kovačević, M., Laalo, K., Palmović, M., Pfeiler, B., Voeikova, M.D., & Dressler, W.U. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Language, 31*, 461-479.

Stem-, Spraak- en Taalpathologie - 27 (Special issue emeritaat Steven Gillis) 2022, pp. 1 - 249

Stem-, Spraak- en Taalpathologie

ISSN 2666-674X