Lexical-semantic deficits in developmental language disorder: the role of statistical learning

Children with developmental language disorder (DLD) have severe difficulty with the acquisition of language. One hypothesis states that these children have a deficit in the ability of statistical learning, a domain-general learning mechanism that is important for extracting patterns and regularities from input implicitly. This dissertation focuses on the potential link between such a deficit in statistical learning and the lexical-semantic difficulties that are experienced by children with DLD. The following research questions are addressed: 1) do children with DLD have a deficit in various statistical learning abilities (word segmentation; cross-situational word learning and semantic categorization) compared to typically developing children and 2) are these statistical learning abilities related to lexical-semantic knowledge in children with DLD? Another research aim was to investigate ways of measuring statistical learning on-line in children with and without DLD.

This dissertation covers four empirical studies. We found evidence for a deficit in cross-situational word learning in children with DLD, indicating that they have more difficulty than typically developing children with learning to couple words to their referents in situations with referential ambiguity. We did not find evidence for or against a deficit in the other types of statistical learning ability in children with DLD, nor did we find evidence for or against a relationship between statistical learning and lexical-semantic knowledge. Concerning our on-line measures of statistical learning, using eye-gaze data as an index of cross-situational word learning shows promise.

ISBN 978-94-6093-421-6
DOI https://dx.medra.org/10.48273/LOT0636

:LOT Netherlands Graduate School of Linguistics
Landelijke Onderzoekschool Taalwetenschap

:LOT 636

Iris Rebecca Lucina Broedelet

Lexical-semantic deficits in developmental language disorder: the role of statistical learning

Iris Rebecca Lucina Broedelet

Lexical-semantic deficits in developmental language disorder: the role of statistical learning

:LOT

University of Amsterdam
Amsterdam Center for Language and Communication

Lexical-semantic deficits in developmental language disorder: the role of statistical learning

Published by
LOT
Binnengasthuisstraat 9
1012 ZA Amsterdam
The Netherlands

phone: +31 20 525 2461

e-mail: lot@uva.nl
http://www.lotschool.nl

Cover illustration: 'Licht in het bos', Delia Alberse, 2022 (detail).

ISBN: 978-94-6093-421-6
DOI: https://dx.medra.org/10.48273/LOT0636
NUR: 616

Lexical-semantic deficits in developmental language disorder: the role of statistical learning

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 8 februari 2023, te 10.00 uur

door Iris Rebecca Lucina Broedelet
geboren te Amsterdam

Promotiecommissie

Promotores:	prof. dr. J.E. Rispens	Universiteit van Amsterdam
	prof. dr. P.P.G. Boersma	Universiteit van Amsterdam
Overige leden:	prof. dr. E.O. Aboh	Universiteit van Amsterdam
	prof. dr. E.H. de Bree	Universiteit Utrecht
	prof. dr. P.J. Monaghan	University of Lancaster
	prof. dr. K. McGregor	University of Iowa
	dr. A.T. Benders	Universiteit van Amsterdam
	dr. M.E.P. Flecken	Universiteit van Amsterdam

Faculteit der Geesteswetenschappen

The research for/publication of this doctoral thesis received financial assistance from the Dutch Research Council (NWO) under Grant 322-89-014.

Voor Elio

Table of contents

Dankwoord

Ik ben ontzettend trots dat het gelukt is om een proefschrift te produceren! Ik wil hiervoor heel graag een aantal mensen bedanken.

Ten eerste wil ik natuurlijk mijn begeleiders Judith en Paul bedanken voor jullie hulp met het schrijven van een onderzoeksvoorstel, het plannen van mijn project, het ontwerpen van mijn taken en het schrijven van mijn proefschrift. Ik vond het een fijne samenwerking!

Judith, al sinds mijn bachelorscriptie (2013) hebben wij veel samengewerkt. Na mijn masterscriptie heb jij me gestimuleerd om een onderzoeksvoorstel te schrijven voor een beurs van NWO. Ik had wat overtuiging nodig, maar ik ben nu heel blij dat we ervoor zijn gegaan! Dankjewel voor je advies, het vertrouwen, de vrijheid die je me gaf, de nodige opkikkers, de nodige kopjes koffie, de gezelligheid en het geloven in mij. Ook je scherpe blik en goede feedback waren heel fijn!

Paul, jou ken ik al sinds ik het vak Fonetische Transcriptie bij jou volgde in 2009. Dit was meteen het leukste vak dat ik ooit gevolgd heb: we moesten namelijk leren alle mogelijke klanken uit te spreken (aβ̞oɰað̞o). Dit vak was echt een openbaring voor mij. Later heb ik heel veel statistiekvaardigheden van jou geleerd. Ik bewonder jouw talent in het begrijpelijk maken van zeer moeilijke en abstracte materie. Je hebt me veel geleerd over kritisch nadenken, iets wat ik niet meer uit kan zetten. Bij elk artikel ga ik na of er niet gedaan wordt aan nulresultaatinterpretatie, *p*-waardevergelijking, of allebei (helaas gebeurt dit heel vaak wel). Dankjewel Paul!

Ik wil ook graag Enoch Aboh, Elise de Bree, Padraic Monaghan, Karla McGregor, Titia Benders en Monique Flecken heel hartelijk bedanken voor het deelnemen aan mijn promotiecommissie. Ik voel me vereerd!

Alle kinderen die hebben meegedaan aan mijn onderzoek: heel erg bedankt! Uiteindelijk is het hele proefschrift gebaseerd op deze data, dus ik ben ontzettend blij dat jullie mee wilden doen aan mijn 'hartstikke leuke' taakjes. Hopelijk viel het niet al te erg tegen. Ik vond het ook heel gezellig om met jullie te kletsen! Ook veel dank aan alle meedenkende ouders, de scholen waar ik langs mocht komen om kinderen te testen, en alle leerkrachten die hun best hebben gedaan om tijd en ruimte voor mij te creëren. Ik ben ook heel blij met de hulp van logopedisten, klinisch linguïsten, ambulant begeleiders, leerkrachten, schooldirecteuren en andere medewerkers van VierTaal, Auris en Pento. Jullie hebben heel veel moeite gedaan om mij te helpen met het werven van kinderen met een TOS en om het verzamelen van data in goede banen te leiden. Voor het verzamelen van de data heb ik ook hulp gehad van testassistenten Jerry, Lia, Maayke, Sonja en Evi. Dankjulliewel! Ook alle volwassen proefpersonen die hebben meegedaan aan mijn pilotstudies, heel erg bedankt!

Ik heb een heel gezellige tijd gehad met mijn collegapromovendi, het voelde bijna een beetje als een tweede studententijd. Dankjewel Imme, Merel, Tessa en Tiffany (6.50 for life!) voor het warme welkom, de gezelligheid en al jullie adviezen. Imme en Merel, met jullie kon ik veel sparren over mijn project en dat was heel fijn. Dankjewel ook Marloes, Marieke, Sybren, Thom, Cindy, Darlene, Ulrika, Natalia, Hernán en Klaas, en mensen die ik nu vergeet (sorry!). Ik ben heel blij dat ik jullie heb leren kennen! Ik vond de lees- en schrijfclub met Judiths promovendi ook heel nuttig en leuk.

Ook wil ik de andere collega's van het ACLC bedanken voor de samenwerking. Dirk Jan Vet, heel erg bedankt voor het programmeren van mijn experimenten, het oplossen van problemen met databestanden en het werkend krijgen van de draagbare eye-trackers. En niet te vergeten de gezellige gesprekken die we daarbij voerden, o.a. hierdoor werd het ontwerpen van experimenten de leukste fase van mijn onderzoek! Ook wil ik hier collega Silke Hamann bedanken. Ik heb tijdens mijn master twee

hele leuke vakken en een tutorial bij jou gevolgd. Alhoewel ik fonologie en fonetiek heel interessant vind, ben ik met mijn PhD een andere kant opgegaan. Onze samenwerking heeft me wel een enorme boost in mijn zelfvertrouwen gegeven, dankjewel daarvoor! Ook een shout-out naar mijn leuke collega's bij het WAP.

Hier wil ik natuurlijk ook mijn allerliefste vrienden bedanken. Sascha, mijn bestie, wij zitten bijna altijd op één lijn en begrijpen elkaar met minimale inspanning. Dankjewel dat je er altijd voor me bent en dat het zo fijn is om met jou tijd door te brengen! Ook een shout-out naar kleine Adinda, en natuurlijk Mark, Judith (Couvee) en Mulan! Wouter, ook jij heel erg bedankt voor je vriendschap, en met zijn drietjes (Sasa Woerdoe en Iri) is het ook altijd zo fijn. Wat hebben we toch enorm veel leuke herinneringen samen! Marcel en de poesjes en Lola, het is ook altijd heel fijn om bij jullie te zijn! Wolf, Elisa en Sinne: we zien elkaar niet supervaak, maar als we elkaar zien is het altijd heel erg fijn en relaxed. Lofjoe! Klaas, Natalia en Hernán, jullie heb ik natuurlijk al als collegae benoemd maar jullie verdienen hier ook een plekje. Zo blij dat ik jullie heb leren kennen, hopelijk blijven we elkaar veel zien! Ook een shout-out naar Elena en Cosmo! Isa, Lisa, Britta, Max, Thomas, Daphne, Harmen, Eva en Rowan, jullie zijn ook lief, bedankt daarvoor!

Sascha en Klaas, heel erg bedankt dat jullie mijn paranimfen willen zijn! Jullie zijn de beste!

Lieve TW'ers Joos, Annelieke, Mart (ɓart), Charlotte, Sheean, Imke en Anne: met jullie begon ik mijn taalwetenschap-carrière! Ik vond onze studententijd echt heel erg leuk samen en ik hoop dat we elkaar blijven zien!

Ik wil ook mijn lieve familie bedanken. Mama, dankjewel voor alles, dat je me met zoveel liefde hebt grootgebracht en me hebt laten voelen dat je trots op me bent. Dat betekent heel veel voor me! Arnoud ook heel erg bedankt voor de steun, je voelt echt als familie. Ook wil ik jullie bedanken voor het oppassen op Elio (handig als je een proefschrift aan het schrijven bent), hij is helemaal dol op jullie! Ook oma (oma Uil), een van de weinigen die dit proefschrift helemaal zal lezen, heel erg bedankt voor je lieve mails en interesse, en onze leuke gesprekken die we al voeren sinds ik kan praten. En natuurlijk de leuke museumbezoekjes met mama! Ook tante Marike, Jannes, Boudewijn en oma Riet wil ik bedanken voor de lieve knuffels, fijne gesprekken en betrokkenheid. Mijn schoonfamilie Zia, Trees en Cees wil ik ook graag bedanken voor jullie interesse in mijn bezigheden. Naast oma Lida wil ik ook de rest van de Alberses bedanken: Mieke, Bever, Gabriella, Sandra, Mees. En de familie van Arnoud: Mariska, Lodewijk, Imme, Max (en Kilian), Arnoud jr. Maaike Bakker: dankjewel dat wij er ook bij horen!

Tom, jij bent mijn rots in de branding, heel erg bedankt voor je steun, liefde, afleiding, alles. Een van de beste beslissingen van mijn leven is om met jou te trouwen. We zijn al zo lang samen maar nog steeds verliefd! Maar belangrijker, het voelt zo veilig en vertrouwd bij jou, en het is zo fijn om samen te zijn, een betere basis om te floreren bestaat niet. Ik ben zo trots op ons en hoe we gegroeid zijn samen, ons gezin, en zo blij met al onze mooie herinneringen. Ik verheug me enorm op onze toekomst. Ik hou van je! Mus en Freddie, ook heel erg bedankt voor jullie zachte aanwezigheid de afgelopen jaren!

Elio, de liefde die ik voor jou voel is niet te beschrijven. Ik ben zo dankbaar voor jouw bestaan. Jouw naam (Elio Anouar) betekent stralende zon, en dat ben jij, je maakt zoveel liefde los in iedereen. Je bent nog te jong om dit proefschrift te begrijpen, alhoewel je al de oren van ons hoofd kletst (dit is TOP als je taalwetenschapper bent: een studieobject!), maar toch draag ik het op aan jou, lieve kleine Elio! Ons gezinnetje is me ongelooflijk dierbaar en maakt me heel erg gelukkig. Ik hou van je!

Author contributions

Iris Broedelet (IB) wrote the proposal for the 'PhDs in the Humanities' funding programme of the Dutch Research Council (NWO). Adjustments were made on the basis of feedback from Judith Rispens (JR), Paul Boersma (PB) and the Faculty of Humanities Grant team of the University of Amsterdam.

Chapter 1 and Chapter 6 were written by IB, with valuable feedback from JR and PB.

Chapter 2 is a slightly modified version of the published article: Broedelet, I., Boersma, P., & Rispens, J. (2021). Measuring (online) word segmentation in adults and children. *Dutch Journal of Applied Linguistics*, *10*. https://doi.org/10.51751/dujal9607. The study was designed by IB, with help from JR and PB. The design of the task was modelled after Haebig et al. (2017) and Gómez et al. (2011). Dirk Jan Vet (DJV) helped with the technical implementation of the experiment and provided test equipment. IB recruited and tested the participants, with help from test assistants Jeremy Bos, Lia Morrissey, Sonja Helder and Maayke Sterk (MS). IB performed data analysis with help from PB and JR. IB wrote the chapter, with feedback from JR and PB. The text was again adapted on the basis of reviewers' comments by IB, in consultation with JR and PB.

Chapter 3 is a slightly modified version of the submitted article: Broedelet, I., Boersma, P., & Rispens, J. (2022). Implicit cross-situational word learning in children with and without developmental language disorder and its relation to lexical-semantic knowledge. *Submitted to Frontiers in Communication: Language Sciences.* The study was designed by IB, with help from JR and PB. The design of the task was modelled after Smith and Yu (2008) and Suanda et al. (2014). The novel words used in the experiment were based on Lammertink et al. (2019), the unusual objects were from Kachergis et al. (2014). DJV helped with the technical implementation of the experiment and provided test equipment. IB recruited and tested the participants, with help from test assistant MS. IB performed data analysis with help from PB. IB wrote the chapter, with feedback from JR and PB.

Chapter 4 is a slightly modified version of the published article: Broedelet, I., Boersma, P., & Rispens, J. (2022). School-aged children learn novel categories on the basis of distributional information. *Frontiers in Psychology*, 12. https://doi.org/10.3389/fpsyg.2021.799241. The study was designed by IB and PB, with help from JR. The design of the task was modelled after Junge et al. (2018) and Chládková et al. (2021). DJV helped with the technical implementation of the experiment and provided test equipment. IB recruited and tested the participants, with help from test assistant MS. IB performed data analysis with help from PB. IB wrote the chapter, with feedback from JR and PB. The text was again adapted on the basis of reviewers' comments by IB, in consultation with JR and PB.

Chapter 5 is a slightly modified version of the submitted article: Broedelet, I., Boersma, P., & Rispens, J. (2022). Distributional learning of novel visual object categories in children with and without developmental language disorder. *Submitted to Language Development Research*. The chapter is based on the same task as is described in Chapter 4. IB recruited and tested the participants, with help from test assistant MS. IB performed data analysis with help from PB and JR. IB wrote the chapter, with feedback from JR and PB. The text was again adapted on the basis of a first round of reviewers' comments by IB, in consultation with JR and PB.

Chapter 1

General introduction

"There was a… caterpillar… green… spinning… red with blue."

A girl with developmental language disorder talks about going to the fair. This sample illustrates the difficulties with expressing meaning that can be caused by this disorder.

Expressing meaning in a precise and sophisticated manner is a human skill par excellence. Children are fantastic word learners, resulting in a mental lexicon that is incredibly extensive and detailed. In the first years of their lives, already starting before reaching six months of age, they learn the meanings of thousands of words. However, the capacity of acquiring language varies between individuals. Some children experience serious difficulties on their path of language acquisition: children with developmental language disorder (DLD; previously known as specific language impairment). Although individual differences within this group are large, children with this disorder can have difficulties in all areas of language, including learning to understand and produce words and all aspects of their meaning. Recent theories have linked DLD to a deficit in statistical learning, a general learning mechanism that underlies the detection of all sorts of patterns and regularities in different types of input. All levels of language contain countless patterns and regularities, and if children with DLD are less capable of detecting these, this could strongly hinder the language acquisition process. This dissertation zooms in on the relationship between statistical learning and lexical-semantic knowledge in children with (and without) DLD. It is investigated whether children with DLD, compared to typically developing (TD) children, have difficulties with different types of statistical learning that could underlie the development of a rich lexicon and whether this ability is related to their existing lexical-semantic knowledge.

1.1 Developmental language disorder and the statistical learning hypothesis

DLD is a developmental disorder that is characterized by a delay and divergence in the acquisition of children native language(s). The diagnosis of the disorder is based on several inclusion and exclusion criteria: the difficulties with language need to be persistent, while they cannot be directly attributed to intellectual disabilities, neurological damage, hearing problems or a lack of language input (Bishop et al., 2017). It is estimated that 7% of school-aged children are diagnosed with the disorder (Bishop, 2006), which would equate to two children in an average Dutch school class. The language difficulties last into adulthood (Botting, 2020) and the disorder has been associated with social-emotional problems (van den Bedem et al., 2018), depression (van den Bedem et al., 2019) and a lower-than-normal quality of life (Eadie et al., 2018).

Numerous different theories about the underlying cause of DLD have been proposed over the years, which can be broadly divided into domain-specific and domain-general theories (Joanisse & Seidenberg, 1998). Domain-specific theories have argued that specifically the linguistic representations are impaired in individuals with DLD, implying that there exists a separate cognitive mechanism for language (van der Lely, 2005). For example, it has been suggested that children with DLD are unable to learn inflectional rules (Gopnik, 1997; Pinker, 1989), that their internal grammar lacks abstract principles of tense and agreement (Rice et al., 1995), or that they have a deficit in the computational system underlying grammar (van der Lely, 2005). On the other hand, domain-general accounts state that deficits in general cognitive abilities, such as a slower than normal general processing speed (Miller et al., 2001), deficits in perceptual processing (Tallal, 1990) or (verbal) working memory (Montgomery et al., 2010), underlie the language difficulties in DLD. Importantly, domain-general accounts also predict that children with DLD show difficulties *outside* of the linguistic domain.

It is the finding of such co-occurring difficulties that has sparked a lot of interest in domain-general theories. It has been reported that individuals with DLD show deficits in, amongst others, working memory

(Montgomery et al., 2010), processing visual information (Collisson et al., 2015), attention (Ebert & Kohnert, 2011) and motor skills (Sanjeevan & Mainela-Arnold, 2019). Recently, a lot of research has focused on the statistical learning deficit hypothesis. Statistical learning is a domain-general learning mechanism that is hypothesized to underlie the extraction of patterns and regularities in all types of input. The statistical deficit hypothesis arises from the idea that language contains endless amounts of patterns and statistical learning mechanisms are thus crucial for language acquisition (Siegelman, 2020). If this learning mechanism is deficient, language acquisition (as well as other abilities) would suffer.

Research into statistical learning surged after Saffran et al. (1996) showed that infants are sensitive to statistical regularities. In their experiment, infants were subjected to an artificial language: a computer-generated auditory stream consisting of four different pseudo-words: (*bidaku, padoti, golabu* and *tupiro*). The words were repeated in a random order and without any pauses between them, resulting in an uninterrupted stream of syllables: …*bidakutupirogolabubidaku*… The aim of the study was to test whether the infants were able to find word boundaries in the stream. Word boundaries were only "marked" by a difference in transitional probabilities between syllables: as *bidaku* was a word, the transitional probability that *da* followed *bi* was 1.0. However, this probability was lower between words: after *ku* either *pa, go* or *tu* could follow, thus the transitional probability between *ku* and either one of these three options was 0.333. Indeed, during the test that was administered after exposure to the stream, the infants listened significantly longer to "part-words" spanning word boundaries (for example *ku-pado*) compared to words like *bidaku,* indicating that infants become sensitive to statistical regularities after only two minutes of passive exposure. Statistical learning is usually viewed as an implicit learning mechanism that does not require explicit instructions or feedback. A later study revealed that adults and children are also able to learn such statistical regularities while doing an unrelated task (Saffran et al., 1997).

Since the study of Saffran et al. (1996), it has been shown that infants, children and adults are sensitive to different types of linguistic and non-linguistic statistical regularities (for a review, see Frost et al., 2019). For example, in a task that is conceptually similar to the word segmentation task, Saffran et al. (1999) showed that infants also become sensitive to patterns in tone sequences. Importantly, statistical learning ability correlates with or predicts language skills, indicating that better statistical learners are better at acquiring different aspects of language (Conway et al., 2010; Ellis et al., 2014; Evans et al., 2022; Gerbrand et al., 2022; Hamrick et al., 2018a; Isbilen et al., 2022; Kaufman et al., 2010; Kautto & Mainela-Arnold, 2022; Kemény & Lukács, 2021; Kidd, 2012; Kidd & Arciuli, 2016; McGregor et al., 2022; Misyak et al., 2010; Newman et al., 2006; Shafto et al., 2012; Spencer et al., 2015; Vlach & DeBrock, 2017).

In children with DLD, deficits in different types of statistical learning have been established (for a review, see Siegelman, 2020). For example, children with DLD perform less accurately on a word segmentation task as described above (Evans et al., 2009; Haebig et al., 2017; Mainela-Arnold & Evans, 2014). Moreover, children with DLD have more difficulty than TD children learning non-adjacent dependencies between novel words (Hsu et al., 2014; Lammertink et al., 2019). Interestingly, they also show deficits on statistical learning tasks that are not linguistic in nature, such as the serial reaction time task, during which participants can implicitly learn motor sequences (Lukács & Kemény, 2014; Lum & Clark, 2022; Mayor-Dubois et al., 2014; Tomblin et al., 2007), visual statistical learning tasks (Collisson et al., 2015; Gillis et al., 2022; Lukács et al., 2021), and learning tone sequences (Ahufinger et al., 2022; Evans et al., 2009). Links between statistical learning ability and language ability have also been reported for children with DLD (Ahufinger et al., 2022; Evans et al., 2009; Hedenius et al., 2011; Mainela-Arnold & Evans, 2014; Misyak et al., 2010; Sack et al., 2021; Tomblin et al., 2007). Please note that null results have been reported (Aguilar & Plante, 2014; Lammertink, Boersma, Rispens, et al., 2020; Noonan, 2018).

Thus, research has shown that statistical learning likely underlies several aspects of language acquisition, and that limitations in statistical learning ability may lead to atypical language acquisition. However, although lexical-semantic deficits occur in children with DLD, not many studies have focussed on the possible connection between a deficit in statistical learning and a divergent development of lexical-semantic knowledge in these children. Therefore, this dissertation focuses on the relationship between statistical learning and lexical-semantic knowledge in children with (and without) DLD. The next sections discuss what kind of lexical problems are associated with DLD and what previous research tells us about the link between statistical learning and lexical-semantic knowledge.

1.2 Lexical-semantic difficulties in DLD

"The major challenge of learning and using a language lies not in the area of broad syntactic principles but in the 'nitty-gritty' of the lexicon."

(Singleton, 1999, p. 4).

As mentioned earlier, children with DLD can have problems in all areas of language, although these problems are heterogeneous in nature. While it might be the case that morphosyntactic difficulties are more prevalent in children with DLD, it cannot be denied that lexical-semantic problems also occur (for a review, see Kan and Windsor, 2010 and Nation, 2014) strongly impacting social and academic development (Aguilar et al., 2017). On average, the first words come later in children with DLD, they have poorer vocabulary breadth and depth than their peers, meaning that they know fewer different words and that the word knowledge that they do have is more superficial (McGregor et al., 2013). They have difficulties finding words and make more semantic (and phonological) errors (Dockrell et al., 2001; Lahey & Edwards, 1999; Leonard et al., 1983; McGregor, 1997; McGregor et al., 2002). Furthermore, children with DLD find it hard to provide sufficient definitions of common words (Dockrell et al., 2003; Dosi et al., 2021; Mainela-Arnold et al., 2010) and their drawings show fewer semantic details than those of TD children

(McGregor et al., 2002; McGregor & Appel, 2002). Their lexical-semantic network also seems to be organized differently, as they have more difficulty producing semantically related words (Drljan & Vuković, 2019; McGregor et al., 2012; Sandgren et al., 2021; Sheng & McGregor, 2010). Children with DLD have more difficulty using the right word in the right context (Charest & Skoczylas, 2019). Learning semantic (and phonological) properties of new words in experiments is also problematic for this group (Alt & Plante, 2006; Haebig et al., 2017; Kan & Windsor, 2010; Nash & Donaldson, 2005), as well as extending novel word meanings (Krzemien et al., 2021). Finally, adolescents with a history of language impairments seem to have difficulty with integrating the meaning of words when processing a sentence, compared to TD peers (Borovsky et al., 2013).

Morphosyntactic errors are regarded as a clinical marker of DLD. At the same time, some researchers view the lexicon as an area of relative strength for children with DLD. In fact, the procedural deficit hypothesis (Ullman, 2016; Ullman & Pierpont, 2005), which is in many ways similar to the statistical learning deficit hypothesis, proposed the idea that there is a dissociation between procedural learning and declarative learning, and that DLD is characterized by a deficit in only the first type of learning and not the latter. According to this hypothesis, procedural learning underlies implicit learning of rule-based motor and cognitive skills and habits, while declarative memory underlies the learning and storage of facts and events. In DLD, a deficit in the procedural memory system thus would affect areas of language that are more structural in nature, such as grammar and phonology, while areas that are more idiosyncratic in nature such as the mental lexicon, are supported by the intact declarative memory system and thus relatively spared. The statistical learning deficit hypothesis on the other hand assumes that regularities in all aspects of language can in principle be learned through statistical learning mechanisms (Hsu & Bishop, 2010). This would predict that a deficit in statistical learning can lead to difficulties in all areas of language. In the next section, what is already known about the role of statistical learning in different stages of lexical development is discussed.

1.3 Statistical learning and the development of lexical knowledge

Research suggests that statistical learning abilities contribute to lexical skills. For example, it has been found that better statistical learning skills correlate with larger vocabularies in children (Spencer et al., 2015). Importantly, a predictive relationship between statistical learning on the one hand and vocabulary on the other hand has also been established in longitudinal studies: visual statistical learning ability and word segmentation ability at a young age predict later vocabulary size (Ellis et al., 2014; Shafto et al., 2012; Singh et al., 2012). In children with DLD, it has also been reported that statistical learning correlates with their vocabulary size (Evans et al., 2009; Mainela-Arnold & Evans, 2014).

In the study of Mainela-Arnold and Evans (2014), the relationship between statistical learning (measured with a word segmentation task) and two types of lexical abilities in children with DLD, lexical-phonological and lexical-semantic skills, is investigated. Lexical-phonological skills were measured using a forward gating task. Children were presented with increasingly larger parts of words and had to guess which word they heard. The number of incorrect guesses was recorded. During the lexical-semantic task, the children had to provide definitions of words. The authors report that statistical learning ability was significantly correlated to lexical-phonological skills. On the other hand, no significant correlation was found between statistical learning ability and lexical-semantic skills. These findings were interpreted as indication of a relationship between statistical learning and sequential lexical-phonological abilities in children with and without DLD, while statistical learning is less likely to contribute to lexical-semantic knowledge. However, this conclusion is based on a *p*-value comparison instead of a direct comparison of the groups (when a correlation between variables is tested in two groups separately, a difference between groups cannot be concluded from different p-values of this correlation; Nieuwenhuis et al., 2011).

This dissertation dives deeper into this relationship by setting up studies in which statistical learning as well as lexical-semantic knowledge are measured more extensively and more directly: three different statistical learning tasks have been constructed that are hypothesized to contribute to different stages in word learning. Passive and active vocabulary size, word category knowledge and lexical-semantic organization are taken into account as measures of lexical-semantic knowledge. At least three different stages of word learning, which likely happen simultaneously and feed into each other, can be distinguished: finding words in the uninterrupted speech stream, mapping word forms to referents in the real world, and organizing new words and concepts into semantic categories. For this dissertation, tasks have been developed that aim to measure the role of statistical learning in these three word-learning processes: a word segmentation task, a cross-situational word learning task and a visual distributional learning task. These tasks target different stages of lexical development and measure different types of underlying statistical structures: sequential, associative and distributional regularities. This is important for a better understanding of the statistical learning abilities in children with DLD. The three types of statistical learning will be explained in more detail below.

1.3.1 Word segmentation

Boundaries between words are not consistently marked by pauses or other prosodic cues. The process of segmenting words from fluent speech is the first word-learning stage that has been hypothesized to be supported by statistical learning mechanisms (Saffran, Aslin, et al., 1996). As stated above, this ability has been linked to vocabulary size in TD children and children with DLD, and it has also been shown that children with DLD perform worse on word segmentation tasks. In this dissertation, a new word segmentation task is developed that is based on the properties of the Dutch language (Chapter 2). Moreover, learning on the task is measured using both off-line and on-line measures (see §1.4).

1.3.2 Word–referent mapping

Linking words to referents is another word-learning process that seems to be supported by statistical learning mechanisms. The task of mapping new word forms to the correct referents is more complicated than one might initially think, if referential ambiguity is considered (Quine, 1960). A child that is learning language, hears a lot of new words and sees a lot of potential referents at the same time. How does the child learn the correct mappings between words and their referents? "In any naming event, a novel word can refer to any object present, its properties, the speaker's feelings or intentions for it, an impending action, or something else altogether" (McMurray et al., 2012, p. 832). Although individual instances are often ambiguous, a certain word occurs with its referent more often than with other objects *across* situations. Statistical learning mechanisms are hypothesized to play a role in resolving referential ambiguity, by (implicitly) tracking the co-occurrences between words and referents (Smith & Yu, 2008). Language learners need exposure to words in different contexts to fully grasp its meaning (see Figure 1.1). Learning word meanings then can be viewed as a gradual, accumulative process.

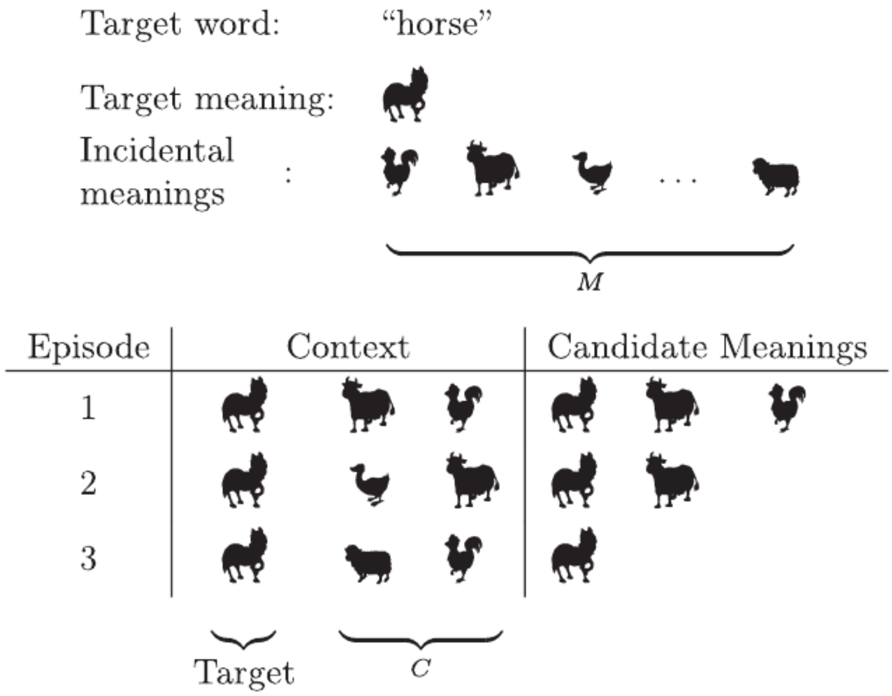


Figure 1.1 – Encountering a word in different contexts helps a child to discover its meaning (picture from Blythe et al., 2010).

Cross-situational word learning tasks are designed to mimic word–referent mapping in situations of referential ambiguity. In the exposure phase of these tasks, participants are subjected to ambiguous learning trials, during which they hear multiple novel words and see multiple unknown referents, without indication of the correct mappings. However, as words occur with the correct referent consistently *across* trials, these mappings can be learned by accumulating evidence over time. Several studies have shown that infants (Smith & Yu, 2008; Vlach & Johnson, 2013; Yu & Smith, 2011), children (Suanda et al., 2014; Vlach & DeBrock, 2017) and adults (Fitneva & Christiansen, 2011; Kachergis et al., 2014; Smith et al., 2011; Suanda & Namy, 2012; Yu & Smith, 2007) are able to learn word–referent pairs after a few minutes of exposure. If children with DLD have more difficulty with picking up these co-occurrences between words and referents, this could hamper their lexical development.

Studies into cross-situational word learning in children with DLD are scarce, but recently, two studies have been published. Ahufinger et al. (2021) showed that children with DLD learn fewer word–referent pairs in a cross-situational word learning task, although both groups performed above chance level. Eye-tracking data did not reveal evidence for on-line learning of word–referent pairs, nor a group difference in on-line learning. Also McGregor et al. (2022) report less accurate cross-situational word learning in children with DLD. They also found that vocabulary was a strong predictor of cross-situational word learning ability. In both studies, the children were explicitly instructed to learn the names of new objects. As learning word meanings is likely not this explicit in actual language acquisition, implicit cross-situational word learning is targeted in this dissertation (Chapter 3). Moreover, to investigate the relationship between cross-situational statistical learning and vocabulary more extensively, several types of lexical-semantic measures are included in our test battery.

1.3.3 Semantic categorization

Besides segmenting words from fluent speech and linking those words to visual referents, language learners also have to divide the world into semantic categories. Semantic categorization could also be supported by statistical learning mechanisms (Unger & Fisher, 2021). Imagine that a child has learned to link the word *dog* to a referent 'dog'. The child then has to learn the limits and specificity of this category. Are all hairy animals with four legs called *dog*? Is only one specific type called *dog*? Is this hairy animal with a long tail a *cat* or a *dog*? Picking up similarities and differences between referents to form semantic categories could be a type of pattern that is learned by statistical learning mechanisms. A well-connected lexical-semantic network is crucial for using words correctly, as semantically related words activate each other. A "poor" network, on the other hand, causes word-finding difficulties and underspecified word use. If children with DLD have difficulty with learning these semantic regularities, this could result in underspecified semantic representations and a less efficiently organized lexicon.

Previous research has indicated that processing semantic information can be supported by statistical learning mechanisms. For example, it is easier for adults to learn implicit mappings between objects in same-category pairs than in different-category pairs (Rogers et al., 2021). More directly related to word learning, statistical learning also seems to contribute to the development of the shape bias, which entails the tendency to generalize words to new objects that share the same shape. Collisson et al. (2015) report that the development of the shape bias lags behind in children with DLD. Moreover, this ability is related to visual statistical learning ability, which is also weaker in children with DLD compared to their peers.

To investigate semantic statistical learning in children with and without DLD, this dissertation also tests distributional learning, which is yet another type of statistical learning (Chapter 4 and Chapter 5). Distributional learning underlies the categorization of different types of stimuli. It was first operationalized in the context of phonological research by Maye et al. (2002, 2008). In their distributional learning task, infants were exposed to speech sounds. The speech sounds were variants from a continuum that ran from voiceless plosive [t] to voiced plosive [d]. The frequency of particular speech sounds either followed a bimodal or a unimodal distribution. In the bimodal distribution, there were two "frequency peaks" while the unimodal distribution had one broader "peak" (see Figure 1.2). In the test phase of Maye et al. (2008), the infants that had been exposed to the bimodal presentation of speech sounds, were significantly better at discriminating stimuli 3 and 6, indicating that the frequency peaks during exposure cause infants to learn that the input contained two distinct categories. Visual versions of this experiment indicate that distributional learning also underlies the categorization of new faces (Altvater-Mackensen et al., 2017) and novel animate objects (Junge et al., 2018). In this dissertation, it is tested if children with DLD have difficulty with visual distributional learning and whether this is related to their lexical-semantic knowledge.

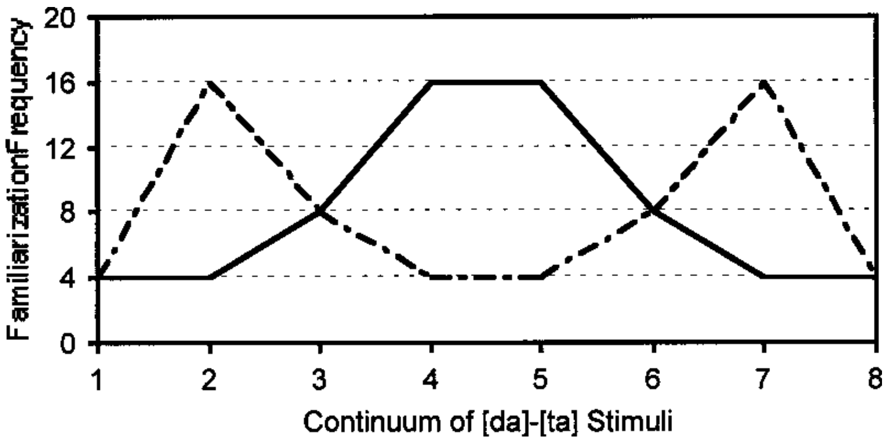


Figure 1.2 – Unimodal versus bimodal distributions in the experiments of Maye et al., 2002, 2008

1.4 Measuring statistical learning ability off-line and on-line

Statistical learning tasks often consist of an exposure phase and a ("off-line") test phase. In the exposure phase, participants are subjected to input containing some type of statistical regularity, for example an artificial language, either passively or while doing an unrelated task. In the test phase, participants' sensitivity to those regularities is measured. The type of test items depends on the task and the age of the participants. As described in §1.1, infants' sensitivity to transitional probabilities can be tested using the preferential looking paradigm. With older children or adults, explicit questions can be used such as "Which word sounds better?". The participant then has to choose between the target word and a distractor word. If participants, on average, choose the target answer significantly more often than would be predicted by chance, this is interpreted as evidence for statistical learning.

However, there are downsides of this way of measuring statistical learning (Arnon, 2020; Siegelman, Bogaerts, & Frost, 2017; Siegelman, Bogaerts, Kronenfeld, et al., 2018). Firstly, it is likely that abilities besides statistical learning impact performance on off-line tasks, such as working memory, encoding skills and meta-linguistic abilities. Moreover, test phases usually consist of at least 16 test items, sometimes a lot more, and stimuli are repeated throughout the test. Exposure to this repetition of targets and distractors, could "overwrite" the (implicit) knowledge that was learned during the exposure phase. Another important point is that mean accuracy on a test phase might not be a reliable indicator of individual differences in statistical learning capacity, as such tasks are originally developed for group-level conclusions. As an important goal of the field is to find links between statistical learning ability on the one hand and language abilities in different participant groups on the other, the development of on-line measures of statistical learning should be encouraged.

Measuring statistical learning on-line means that learning is measured already during exposure to the statistical regularities. Different types of behavioural or neurophysiological methods can be used to measure learning, for example recording reaction time to certain stimuli, eye-movements or brain activity. In contrast to off-line measures, on-line measures have the potential to provide insight into the learning trajectory as opposed to solely the "end product" of statistical learning. Moreover, on-line measures might be more sensitive to (individual differences in) statistical learning as they likely tap better into newly developing knowledge of statistical regularities. This is especially important when testing statistical learning in children. Off-line test phases are very challenging for them, and even more so for children with DLD. In this dissertation, therefore, an aim was to test methods for measuring statistical learning on-line in school-aged children with and without DLD.

1.5 Research questions and dissertation outline

This dissertation aims to investigate 1) whether children with DLD have a deficit in statistical learning abilities (word segmentation; cross-situational word learning and semantic categorization) compared to TD children and 2) whether these statistical learning abilities are related to existing lexical-semantic knowledge in children with and without DLD. In other words, does a statistical learning deficit contribute to the lexical-semantic impairments in children with DLD? Another aim of this dissertation is 3) investigating ways of measuring statistical learning on-line in children with and without DLD. With these aims in mind, three tasks that measure different types of statistical learning were constructed: a word segmentation task, a cross-situational word learning task and a visual distributional learning task. These tasks are discussed in different chapters in the dissertation. Chapter 2 addresses word segmentation in children with and without DLD. The chapter reports results of a word segmentation task in which word boundaries can only be learned on the basis of transitional probabilities. Moreover, a method of measuring statistical learning on-line by recording participants' reaction times when responding to click sounds during the exposure phase of the task (Franco et al., 2015; D. M. Gómez et al., 2011) is tested in children for the first time. It is investigated whether children with DLD show less efficient on-line learning and off-line performance compared to TD children.

Chapter 3 reports a study on implicit cross-situational word learning in children with and without DLD. During this task, children can learn word–referent pairs by tracking the co-occurrences between words and referents across trials. The children are not explicitly instructed and perform a non-related task during the exposure phase. As an on-line measure of word–referent learning, eye-movements are recorded. Afterwards, knowledge of the word–referent pairs is tested using a multiple-choice task. TD children and children with DLD are compared on off-line test results and on-line eye-tracking data, as well as the relationship between cross-situational word learning ability and several lexical-semantic measures. Chapter 4 reports a newly designed task which measures visual distributional learning which is suitable for school-aged children. The task is based on Junge et al. (2018) and measures whether exposure to distributional information impacts the categorization of novel animate stimuli. This study is innovative because it tests visual distributional learning in school-aged children for the first time. Moreover, a possible confound in earlier distributional learning task is avoided by using a novel design (Chládková et al., 2022). Chapter 5 compares visual distributional learning ability (measured using the task that is described in Chapter 4) between children with and without DLD, and investigates whether this ability is related to lexical-semantic knowledge. Finally, Chapter 6 summarizes the results and addresses the research aims that are stated in the beginning of this section.

1.6 Data availability statement

For all studies that are reported in this dissertation, data and scripts used for analysis are openly available on FigShare:
https://figshare.com/authors/I_R_L_Broedelet/4481404.

Chapter 2

Measuring (on-line) word segmentation in adults and children

This chapter is a slightly modified version of the published article:

Broedelet, I., Boersma, P., & Rispens, J. (2021). Measuring (online) word segmentation in adults and children. *Dutch Journal of Applied Linguistics*, *10*. https://doi.org/10.51751/dujal9607

R scripts, data and materials are available on FigShare: https://figshare.com/collections/_/4739162

Abstract

Since Saffran et al. (1996) showed that infants were sensitive to transitional probabilities between syllables after being exposed to a few minutes of fluent speech, there has been ample research on statistical learning. Word segmentation studies usually test learning by making use of "off-line methods" such as forced-choice tasks. However, cognitive factors besides statistical learning, such as encoding, memory and meta-linguistic abilities likely influence performance on those tasks. The goal of the present study was to improve a method for measuring word segmentation on-line. Click sounds were added to the speech stream, both between words and within words. Stronger expectations for the next syllable within words as opposed to between words were expected to result in slower detection of clicks within words, revealing sensitivity to word boundaries. Unexpectedly, we did not find evidence for learning in multiple groups of adults and child participants. We discuss possible methodological factors that could have influenced our results.

2.1 Introduction

Language is full of patterns and regularities. In the last decades, there has been great interest in the role of *statistical learning* in language acquisition. Statistical learning is a cognitive ability that underlies the implicit discovery of statistical patterns and sequences in sensory input (Siegelman, Bogaerts, Kronenfeld, et al., 2018) and has been hypothesized to contribute to different areas of language acquisition (for a review, see Romberg and Saffran, 2010). One of the first demonstrations of statistical learning was the seminal study of Saffran et al. (1996). As word boundaries are not (consistently) marked by pauses or other prosodic cues in natural speech (Cole, 1980), the authors aimed to investigate whether statistical learning plays a role in learning to recognize separate words in a stream of speech sounds. Eight-month-old infants with English-speaking parents were exposed to a two-minute synthesized stream of uninterrupted syllables. The speech stream consisted of four pseudo-words (*bidaku*, *padoti*, *golabu* and *tupiro*) that were repeated in a random order. The authors wanted to test whether infants were able to recognize these pseudo-words after exposure to the stream, despite the absence of any prosodic cues for word boundaries. Results from a head-turn preference procedure administered after familiarization show that infants listen longer to "part-words" that span word boundaries, such as *ku-pado*, than to target words. This novelty preference indicates that infants learn to recognize target words and are thus sensitive to the statistical probabilities of the input: the transitional probabilities (TPs) between syllables. For example, the probability that *da* followed *bi* in the stream was 1.0, while the probability that *pa* followed *ku* was only 0.333. As there were no pauses or other prosodic cues for word boundaries[1], infants' learning could only have been happened based on these TP values.

The degree of learning in statistical learning tasks is usually inferred from participants' performance on an "off-line" task which they undergo after the familiarization phase, during which they have to choose between

1 Prosody does play an important role in word segmentation (see for example Endress & Hauser, 2010).

target words and foils. However, performance on such tasks could be strongly influenced by cognitive processes other than statistical learning, such as encoding and memory capacities, meta-linguistic skills and decision-making biases (Siegelman, Bogaerts, Kronenfeld, et al., 2018). Specifically for children, meta-linguistic questions such as "which word sounds better?" are difficult to process and answer, which could lead to underestimation of their (implicit) knowledge. Importantly, while statistical learning is a continuous process, off-line measures provide information about behaviour at only a single point in time. On-line methods, on the other hand, can provide more insight into the trajectory of statistical learning by measuring learning throughout the familiarization phase. It has thus been argued that in future statistical learning studies, especially those focusing on children, it is important to develop sensitive on-line measures of statistical learning (Lammertink et al., 2017; Siegelman, Bogaerts, Kronenfeld, et al., 2018).

Based on the idea that reaction time reflects processing time, stimulus detection tasks have been used as on-line measures of sentence processing (e.g. Cohen & Mehler, 1996; Fodor & Bever, 1965; Foss & Lynch, 1969), and this paradigm has also been applied to word segmentation tasks. Gómez et al. (2011) added click sounds to the speech stream in their word segmentation task. Italian-speaking adults were asked to listen to a stream of speech sounds, consisting of four pseudo-words (*pabuda*, *gifoto*, *venola* and *minaro*) for four minutes and to push a button as fast as possible when they heard a click sound. Crucially, the clicks occurred either *between* two pseudo-words or *within* a pseudo-word (compare *pabuda*!*gifoto* to *pa*!*buda,* where ! indicates a click). The authors hypothesized that participants who had learned to recognize words in the stream should have stronger expectations for the next syllable when hearing the first syllable of a word compared to when they hear the final syllable of a word. This, in turn, should lead to a larger surprise effect (and thus a slower reaction time) when detecting clicks occurring within words compared to clicks between words. Results showed that after two minutes of exposure, people are indeed slower when detecting clicks within words as opposed to clicks between words, indicating sensitivity to word boundaries that develops over time.

Franco et al. (2015) aimed to replicate these results and tested French-speaking adults on a similar task. As opposed to Gómez et al. (2011), the researchers did not find evidence for a difference in response times to clicks between words and clicks within words. Ten out of 28 participants showed the expected pattern while the other 18 showed the opposite pattern. In their second experiment, Franco et al. (2015) compared performance on two versions of the task: a "passive" word segmentation task with clicks to which participants did not have to respond ("passive-click"); and a word segmentation task without any clicks ("no-click"). They found that performance on the off-line test phase of the passive-click version was significantly lower than performance on the no-click condition, indicating that the statistical learning process might have suffered from the addition of clicks to the stream. Hearing the clicks might have diverted attention from the syllable structure in the input, as participants might have focused more on detecting the clicks than on the artificial language. Another possibility is that the clicks might have given participants false cues to word boundaries, as the clicks were the only "prosodic" elements in the speech stream. The click detection paradigm has the potential to reveal the word segmentation process minute by minute[2], but the finding of mixed results might indicate that an adaptation of the paradigm is called for.

2.2 The current study

Our aim was to find a method for measuring word segmentation on-line that would be suitable for adults as well as for children. As Gómez et al. (2011) and Franco et al. (2015) found mixed results, we decided to adapt the click detection by extending the familiarization phase to eight minutes to facilitate learning. The first and final two minutes contained only a few

[2] While neurophysiological measures like EEG offer an excellent temporal resolution (see for example Kooijman et al. (2005), these methods are costly and more difficult to carry out with children compared to behavioural methods.

click sounds and were added to provide the participants with more "clean" input (without potential distraction from clicks) to facilitate learning of word boundaries. Based on previous studies, we hypothesized that participants could use statistical information to segment words from uninterrupted speech and that our adaptations to the task would result in a learning effect: slower reaction times for clicks within words compared to clicks between words. We constructed an artificial language based on the study of Haebig et al. (2017), as they tested a similar participant group as we intend to test for our future studies (school-aged TD children and children with developmental language disorder; DLD). We conducted three separate experiments. In our first experiment we tested on-line word segmentation using the click detection task. As we did not find evidence for learning on either the click detection nor the off-line task, we conducted a second experiment in which we removed the click sounds to test whether participants (adults and children) would show learning on the off-line task. Finally, as we did not find evidence for off-line learning in Experiment 2, we conducted Experiment 3 in which we used non-words (TP = 0) as foils instead of part-words (TP = 0.333), to test whether adults would learn to distinguish words from non-words.

2.3 Experiment 1

2.3.1 Methods and materials

Participants. Thirty-one adults (21 females and 10 males) participated in the study. Their ages varied between 19;8 (years;months) and 35;11 (*M* = 28;4, *SD* = 6;4). All participants were native speakers of Dutch and had been brought up monolingually. The participants reported that they did not have any hearing difficulty, serious visual problems, developmental dyslexia or any other language-based disorders, ADHD, ASD or learning difficulties. People who (had) studied linguistics or had taken courses in linguistics were excluded from participation. Ethical approval for the experiment was obtained from the Ethical Committee of the faculty of Humanities of the University of Amsterdam. All participants filled in an informed consent form.

Stimuli and design: familiarization phase. We constructed a speech stream from recorded and modified speech. Two sets of four bisyllabic words were constructed to control for order effects: /kiba/, /moti/, /dalu/, /χido/ (language A) and /bamo/, /tida/, /luχi/, /doki/ (language B). There was no significant difference in mean phonotactic frequency in Dutch between the words of language A (M = 1.425, SD = 0.174) and the words of language B (M = 1.385, SD = 0.189): t[3] = 0.738, p = 0.37). All syllables were recorded by a female native speaker of Dutch in a soundproof room. To ensure natural co-articulation between all syllables in the stream, three-syllable sequences were recorded of which the middle syllable was used to construct the stream (see Table 2.1). For example, to construct part of the stream *lukiba*, we recorded *da**lu**ki*, *lu**ki**ba* and *ki**ba**mo* and used the middle syllables (see Graf Estes, 2012). All sound editing was done using the software Praat (Boersma & Weenink, 2019).

Table 2.1 - Three-syllable sequences that were recorded for language A and language B. The bolded letters represent the syllables that were used to construct the stream.

Language A				Language B			
ki	ti**ki**ba	lu**ki**ba	do**ki**ba	*ba*	da**ba**mo	χi**ba**mo	ki**ba**mo
ba	ki**ba**mo	ki**ba**da	ki**ba**χi	*mo*	ba**mo**ti	ba**mo**lu	ba**mo**do
mo	ba**mo**ti	lu**mo**ti	χi**mo**ti	*ti*	mo**ti**da	χi**ti**da	ki**ti**da
ti	mo**ti**ki	mo**ti**da	mo**ti**χi	*da*	ti**da**ba	ti**da**lu	ti**da**do
da	ba**da**lu	ti**da**lu	do**da**lu	*lu*	mo**lu**χi	da**lu**χi	ki**lu**χi
lu	da**lu**ki	da**lu**mo	da**lu**χi	*χi*	lu**χi**ba	lu**χi**ti	lu**χi**do
χi	ba**χi**do	ti**χi**do	lu**χi**do	*do*	mo**do**ki	da**do**ki	χi**do**ki
do	χi**do**ki	χi**do**mo	χi**do**da	*ki*	do**ki**ba	do**ki**ti	do**ki**lu

A unique 8-minute pseudo-random sequence of the four words was generated for each participant, with the restriction that a word could not occur twice in a row. Transitional probabilities between syllables were high within a word (TP = 1). For example, /ba/ always followed /ki/ in language A. Across word boundaries, transitional probabilities were lower, as for example /ba/ could be followed by either /mo/, /da/ or /χi/ (TP = 0.333) in language A. The stream was constructed such that there were no pauses or other prosodic cues for word boundaries: speech was monotone and all syllables were equally long (consonants 118 ms and vowels 160 ms). The syllable rate was 216 syllables (108 words) per minute, resulting in a total of 864 words per participant, with each of the four words occurring 216 times. The stream started with the second syllable of a word and ended with the first syllable of a word, so that the stream did not start or end with a word boundary.

High-pitched 20 ms click sounds (created in Praat) were inserted at random positions in the stream for each participant. There were always at least four syllables in between two clicks, to make sure participants had enough time to respond to every click. Importantly, half of the clicks occurred between two words (for example *kiba*!*dalu*) while the other half were placed within a word (for example *da*!*lu*). The clicks were 1.6 times louder compared to the speech sounds to facilitate the detection of the clicks. The first and final parts of the familiarization phase (2 minutes each) contained 10 clicks, while the middle part (4 minutes) contained 72 clicks. The practice block (30 s), which was included to get the participants used to the click detection task, contained 5 to 6 clicks (see Table 2.2).

Table 2.2 – The structure of the familiarization phase of the word segmentation task.

	Practice	Part 1	Part 2	Part 3
Duration of block	30 s	2 min	4 min	2 min
Total nr. of clicks	5-6	10	72	10
Clicks per minute	10	5	18	5
Percentage clicks	20%	4%	16%	4%

Stimuli and design: off-line test phase. The off-line test phase consisted of 16 two-alternative forced-choice items, in which the targets were words (for example *kiba* for language A) and foils were part-words (syllable combinations spanning word boundaries, for example *bamo* for language A). The four targets were combined with each of the four foils to construct 16 test items. The targets of language A were used as foils for language B and vice versa. The test items were recorded in a citation form by the same female speaker who recorded the stimuli for the familiarization phase and were edited in the same way as the sounds used in the familiarization phase. The order of the test items was randomized for each participant with the restriction that stimuli (either as target or foil) could not appear in two test items in a row.

Procedure. The experiment was carried out in a quiet room in the speech lab of the University of Amsterdam. The experiment was executed in E-Prime 2.0 software (Psychology Software Tools Inc, 2002). Participants sat behind a laptop computer screen wearing headphones and holding a response box. Test version was counterbalanced across participants. Pre-recorded child-directed[3] instructions told them to carefully listen to "a weird language", to press the button as fast as possible when they heard a click sound, and to pay attention as there would be questions at the end. Participants first practiced the click detection task for 30 seconds and proceeded on to the familiarization phase when confirmed they understood the task. As visual feedback, a hashtag (#) appeared on the screen when the button was pushed. In the test phase, participants heard two sequences for every test item, and were asked to choose which one sounded the most like the language they had just heard. There was one practice item. The numbers 1 and 2 appeared on the screen and the participants had to use the two corresponding buttons on the response box. It was possible to repeat test items once. All participants did another statistical learning task as well, the results of which are not discussed in this chapter. Testing took approximately 30 minutes per subject and everyone was compensated with 5 euros for their participation.

Analysis. Data was analysed using the free software R (R Core Team, 2020). For the off-line measure, the practice test item was excluded from further analysis. To compute accuracy, test items were scored as correct when the participant chose the target word, and as incorrect when the participant chose the foil. For the on-line measure, only responses to the clicks from the second block were taken into account (72 clicks per participant). A response was considered valid when it occurred within 2

[3] The task was developed in such a way that it should be suitable for child participants as well, as we intended to test on-line word segmentation in children in a later stage of the project.

seconds after a click. Missed clicks and extraneous responses were removed from the data (1.64%). One participant was excluded from analysis due to too many missed clicks (38) and extraneous responses (27). This resulted in data suitable for analysis from 30 participants. As the RT data were not normally distributed (see Figure 2.1), they were normalized for further analysis to meet the normality assumption of mixed effect models: the response times were first ranked from 1 to *N* (where *N* is the total number of observations) and then normalized using *qnorm*((*rank* – 0.5)/*N*) in R.

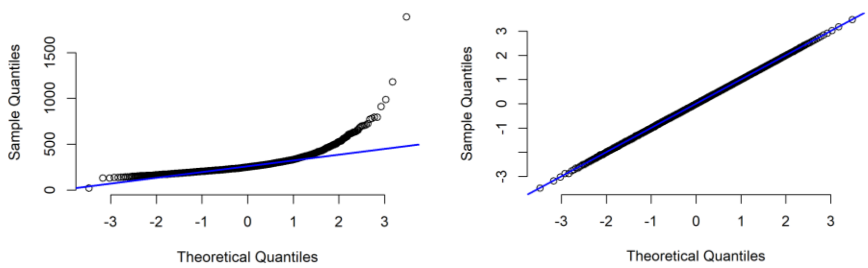


Figure 2.1 – Distribution of the RT data before and after normalization.

2.3.2 Results

Off-line test phase. The average accuracy on the off-line test phase was 0.45 (*SD* = 0.17). A generalized logistic linear mixed-effects model (from the package lme4: Bates et al., 2015) was constructed to test whether participants performed above chance level (0.50). The dependent variable was Accuracy (a 1 or 0 value for every item). Between-participant predictors were Version (A/B) and TargetOrder (first/second; meaning whether the target was heard first or second during a particular test item). The different levels of the predictors were coded into sum-to-zero orthogonal contrasts (Kraemer & Blasey, 2004): Version was coded as -1/2 for A and +1/2 for B, and TargetOrder was coded as -1/2 for first and +1/2 for second. We implemented random intercepts by Participant and by Item, as well as by-participant random slopes for TargetOrder and by-item random slopes for Version.

The estimate for the intercept (converted into probability) was 0.42 (95% CI: 0.34 … 0.50). This performance is significantly below chance level (z = -2.016, p = 0.044), from which we might conclude that Dutch adults prefer part-words over words in the off-line test phase of the current word segmentation task. This result is contrary to our expectations and, being one of our exploratory results, may be a chance finding. The effects of Version and TargetOrder on response times were not significant. See Figure 2.2 for the descriptive accuracy data and Table 2.3 for the results of the model.

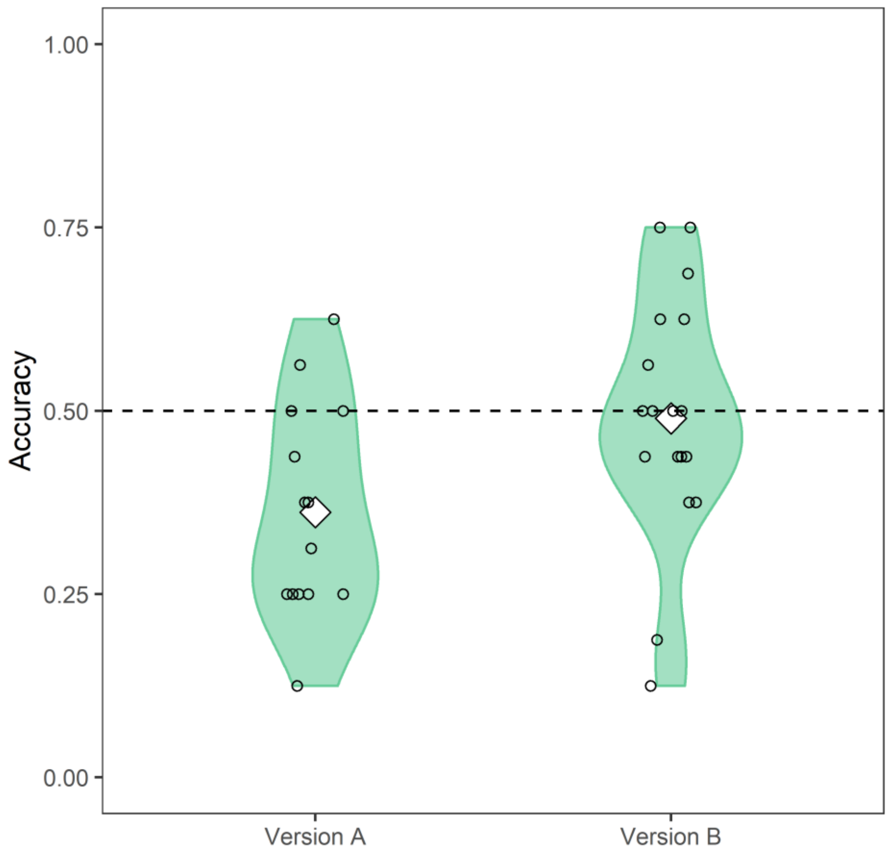


Figure 2.2 – Descriptive plot of participants' accuracy in version A and B of Experiment 1.

Table 2.3 – Results from the linear mixed-effects model.

	Intercept	Version	TargetOrder
Estimate	Probability: 0.42	Odds: 1.72	Odds: 1.40
95% CI	0.34 … 0.50	0.97 … 3.07	0.80 … 2.60
z	-2.016	1.94	1.013
p	0.044	0.052	0.311

Click detection task. A linear mixed-effects model was conducted to test whether the position of the clicks (ClickPosition) influenced their processing time. The dependent variable was normalized RT. Within-participant predictors were ClickPosition (within words/between words) and Block (the middle part of the familiarization phase was divided in four blocks of 1 minute, each containing 18 clicks). Version (A/B) was a between-participant predictor. We implemented random intercepts by Participant and by Item, as well as by-participant random slopes for ClickPosition and Block and by-item random slopes for Version. The different levels of the predictors ClickPosition and Version were coded into sum-to-zero orthogonal contrasts: ClickPosition was coded as -1/2 for between words and as +1/2 for within words, and Version was coded as -1/2 for A and +1/2 for B. The factor Block (1–4) was centred by subtracting 2.5 (the mean), resulting in the following numbers for the four blocks: -1.5, -0.5, +0.5 and +1.5. We expected that an on-line learning effect should surface as a main effect of ClickPosition and/or an interaction between ClickPosition and Block, the latter meaning that RTs are influenced by their context and that this difference is influenced by the amount of exposure to the stream of speech sounds. For the descriptive data, see Table 2.4 and Figure 2.3.

Table 2.4 – Descriptive data: raw and normalized response times for the click detection task.

Raw	Block 1	Block 2	Block 3	Block 4	Overall
Overall RT	275 ms	280 ms	277 ms	290 ms	280 ms
RT between words	271 ms	276 ms	275 ms	293 ms	278 ms
RT within words	278 ms	285 ms	278 ms	288 ms	282 ms
Normalized	**Block 1**	**Block 2**	**Block 3**	**Block 4**	**Overall**
Overall RT	-0.0096	-0.0073	-0.0647	0.0817	0
RT between words	-0.0022	-0.5109	-0.0289	0.0916	0.0012
RT within words	-0.1681	0.0351	-0.1055	0.0726	-0.0012

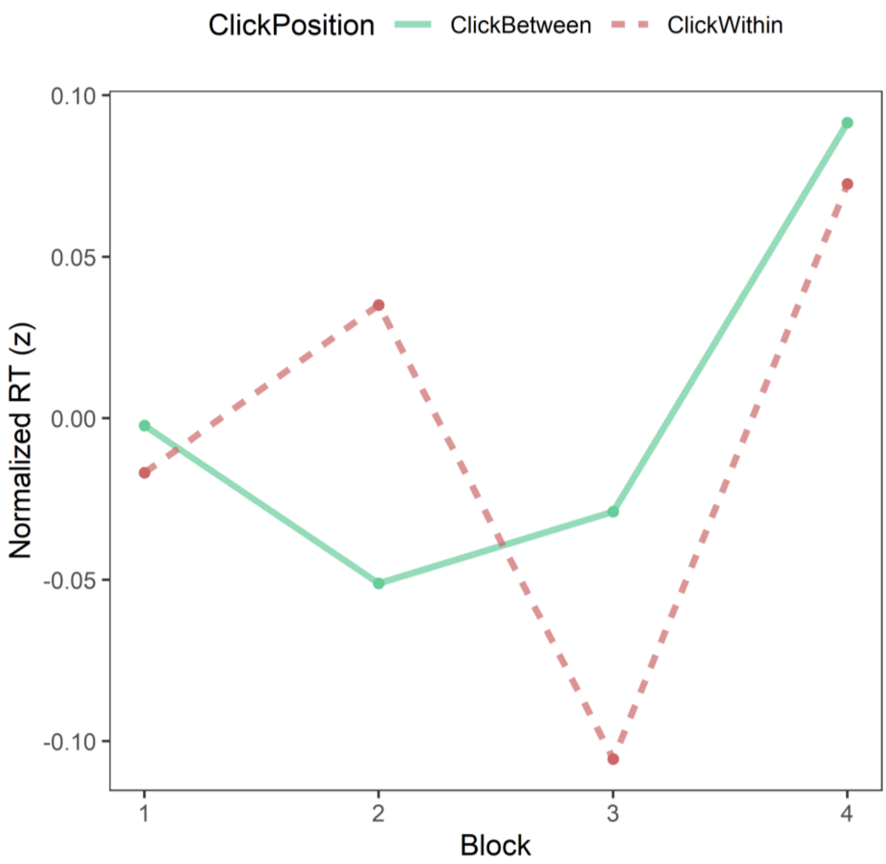


Figure 2.3 – Normalized RT data (z scores) Experiment 1.

The main effect of ClickPosition (estimated Δz = 0.001, 95% CI -0.087 … +0.084) was not significant: t = -0.031, p = 0.98. Neither was the main effect of Block (estimated Δz = 0.024, 95% CI -0.025 … +0.072): t = 0.917, p = 0.35. The interaction between ClickPosition and Block (estimated $\Delta\Delta z$ = -0.04, 95% CI -0.11 … +0.03) also was not significant: t = -1.090, p = 0.26. On the basis of these results, we cannot conclude whether the position of a click (between words or within a word) influenced their processing time, i.e. whether the click detection task revealed sensitivity to word boundaries in the word segmentation task.

There was a significant three-way interaction between ClickPosition, Block and Version (estimated $\Delta\Delta\Delta z$ = -0.16, 95% CI -0.30 … -0.18): t = -2.163, p = 0.036, indicating that the effect of ClickPosition is modified by Block and Version. This is illustrated in Figure 2.4. For version A, the effect of ClickPosition developed as expected from Block 2 onwards and increased over time. For version B, however, the effect reversed in the third block. Individual data (Figure 2.5) shows that there was a large amount of variation in the effect of ClickPosition between participants. Some participants showed a difference in the expected direction, while others showed (almost) no difference or even a difference in the opposite direction.

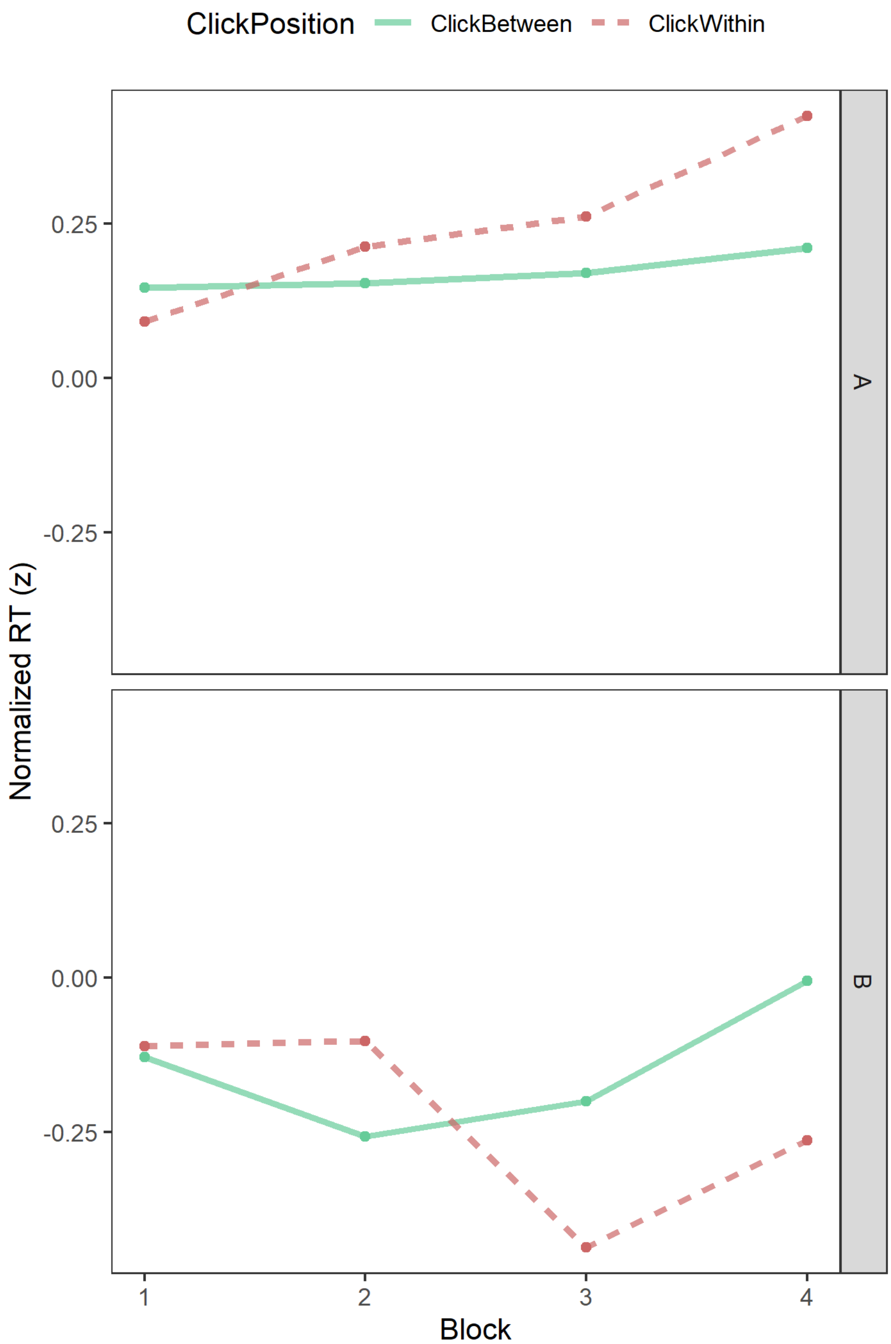


Figure 2.4 – Normalized RT data Experiment 1: version A vs. version B.

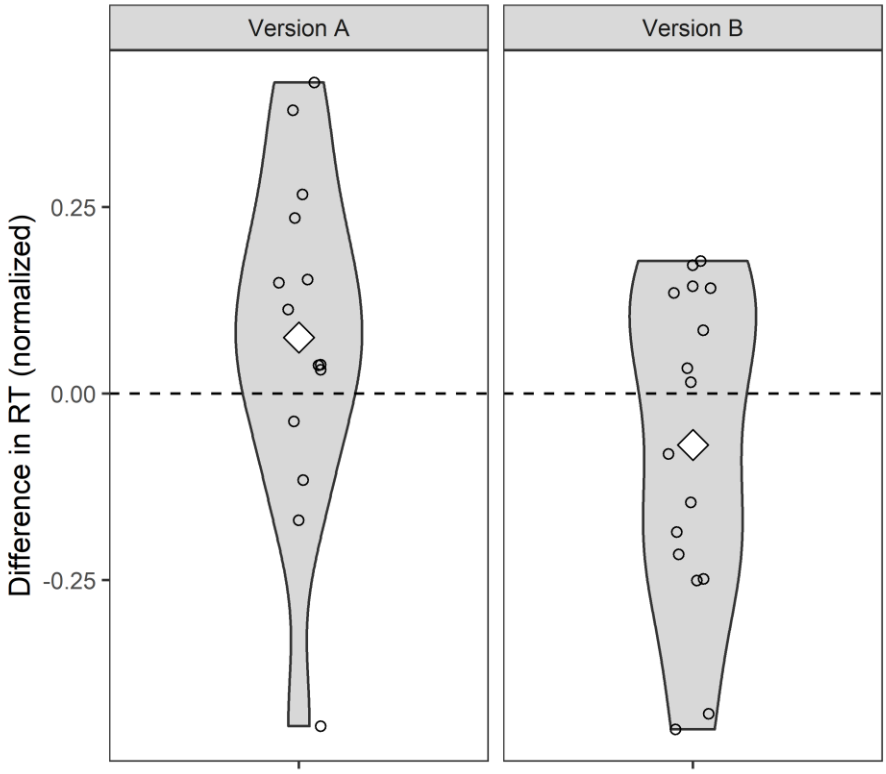


Figure 2.5 – Individual RT data Experiment 1: Mean difference in RT for clicks between words and clicks within words. A positive value implies a learning effect.

2.3.3 Discussion

The aim of Experiment 1 was to adapt the click detection paradigm such that it would be a suitable method to measure word segmentation on-line. As we did not find evidence for or against an on-line learning effect, our results do not support the findings of Gómez et al. (2011). The extension of the familiarization phase does not seem to have been helpful for improving the click detection task as an on-line measure of statistical learning. Exploratorily, we found an unexpected difference between the two test versions: the RTs of the participants who did version A of the task showed the course that we expected, but participants who did version B showed a different pattern.[4] Moreover, similar to Franco et al. (2015), we observed a large amount of individual variation between participants. It could be the case that the words of version A are somehow "easier" to learn, but it might also be true that the click detection task as an on-line measure of learning works for some people (in the way we expect), but not for all.

The fact that we did not find evidence for (or against) learning may reflect that click detection is not suitable as an on-line method, or that it actually negatively influences statistical learning. As Franco et al. (2015) found that performance on an off-line test phase was better when participants listened to a stream without click sounds than when they listened to a stream with click sounds, the authors suggested that the addition of the click detection task, or even just the click sounds, might have hampered statistical learning. Our result of below-chance performance on the off-line test phase could be a chance finding, but it could also be the case that the addition of an on-line measure negatively affected performance on the off-line test phase (Toro et al., 2005). Another explanation might be that the click sounds gave participants false cues for word boundaries. As we cannot draw any conclusions on the basis of only this experiment, we conducted another experiment in which we tested two new groups of participants on the same word segmentation

[4] A reviewer put forward the interesting suggestion that the different performance we found for the two versions might be an item effect, as it could be the case that participants responded faster to clicks occurring in syllables that are more regular in Dutch. If we compare the mean phonotactic probability of the syllable sets *ba*, *ti*, *lu*, *do* (M = 1.96) vs *ki*, *mo*, *da*, *χi* (M = 1.656), which either contain between-word clicks or within-word clicks depending on the test version, we do not find a significant difference: t = -1.295, p = 0.243. Although the difference is not significant, the direction of the difference does correspond to the pattern that responses to clicks across blocks in language A developed more according to our expectations compared to language B. For language A, syllables that contained between-word clicks words had a higher phonotactic probability in Dutch than syllables that contained within-word clicks while it was the other way around for language B. As we hypothesized that between-word clicks should have been detected faster, this difference could have contributed to the finding that the results for language A were more as we expected than the results for language B.

task without the addition of the click detection task. As our intended participant group for future studies on (on-line) word segmentation are school-aged children with and without developmental language disorder (DLD), we included a group of adults and a group of school-aged children in our second experiment.

2.4 Experiment 2

2.4.1 Methods and materials

Participants. Thirty adults (22 female, 8 male) participated in the study. Their ages varied between 18;0 and 26;10 (*M* = 20;6, *SD* = 2;4). Moreover, 30 children (20 girls, 10 boys) between the ages of 7;10 and 10;0 (*M* = 8;5, *SD* = 0;7) participated in the study. The data of three children (one girl, two boys) were excluded because of a diagnosis of developmental dyslexia (N=2) or not speaking Dutch as a native language (N=1), resulting in 27 child participants. The caretakers of the child participants gave active written consent for their participation.

Design. The familiarization phase and off-line test phase were the same as in Experiment 1, except that the click sounds were not inserted into the stream.

Procedure. The experiment was carried out in a quiet room in the speech lab of the University of Amsterdam (adults) or a quiet room in the school of the children. The procedure was the same as in the previous experiment, except that during the familiarization phase, participants were asked to colour a mandala (adults) or a colouring page (children), similar to the study by Saffran et al. (1997)[5]. Testing took approximately 15 minutes. Participants received 5 euros (adults) or sticker sheets (children) as compensation for their participation.

---

[5] In a pilot study, participants did the familiarization phase without any other task but listening to the language. Participants reported that 8 minutes seemed very long and that they felt uneasy.

2.4.2 Results

The mean accuracy on the off-line test phase was 0.49 (*SD* = 0.20) for adults and 0.51 (*SD* = 0.13) for children. See Figure 2.6 for the descriptive data. Two generalized logistic linear mixed-effects models (see section 3.2.1) were conducted. For the adults, the estimate for the intercept (converted into probability) was 0.50 (95% CI: 0.41 … 0.59), which is not significantly different from chance level (z = -0.123, p = 0.90). The main effect of Version was not significant. There was a significant effect of TargetOrder: the odds that an item in which the target was played first was answered correctly were 1.66 (95% CI: 1.02 … 2.46) times higher compared to an item in which the foil was played first: z = 2.210, p = 0.027. For the children, the estimate for the intercept (converted into probability) was 0.51 (95% CI: 0.45 … 0.56), which is not significantly different from chance level (z = 0.240, p = 0.81). The main effects of Version and TargetOrder were not significant. See Table 2.5 for the results of the model. Based on these null results, we cannot conclude whether adults and/or children do or do not distinguish words from part-words, indicating knowledge of word boundaries, in the off-line test phase of the word segmentation task without the click detection task.

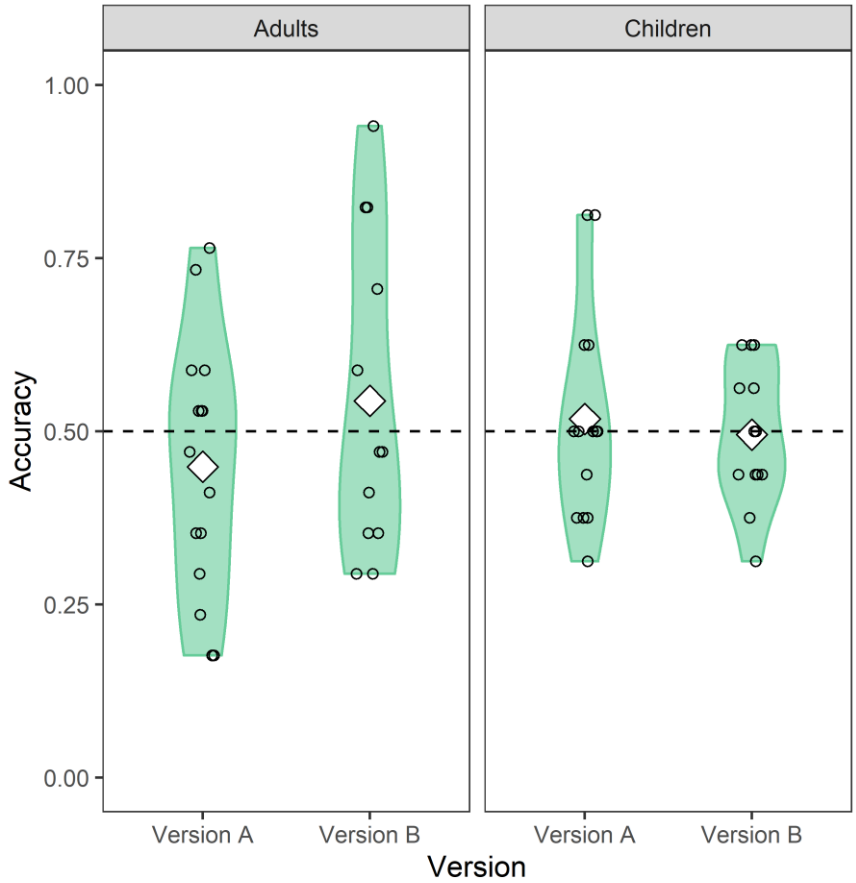


Figure 2.6 – Descriptive plot of adults' and children's' accuracy in version A and B of Experiment 2.

Table 2.5 – Results from the linear mixed-effects model.

	Intercept		Version	
	Adults	*Children*	*Adults*	*Children*
Estimate	Probability: 0.50	Probability: 0.51	Odds: 1.52	Odds: 0.91
95% CI	0.41 … 0.59	0.45 … 0.56	0.52 … 4.51	0.43 … 1.90
z	-0.123	0.240	0.810	-0.273
p	0.90	0.81	0.42	0.79

	TargetOrder	
	Adults	*Children*
Estimate	Odds: 1.66	Odds: 1.22
95% CI	1.02 … 2.46	0.75 … 2.01
z	2.210	0.839
p	0.027	0.40

2.4.3 Discussion

We did not find evidence for (or against) sensitivity to word boundaries in a group of adults and a group of school-aged children in our second experiment, in which we removed the click detection task from the word segmentation task, let alone that we could have anything to say about whether the null result in Experiment 1 was due to interference of the click detection task. For the adults there was a significant effect of the order of the targets and foils in the test items, indicating that it is easier to recognize a target when it is played first in a test item. This factor should be considered when analysing 2AFC data.

Our result is unexpected, as previous studies on word segmentation tasks (Batterink & Paller, 2019; Evans et al., 2009; Finn et al., 2018; Franco et al., 2015; Haebig et al., 2017; Mainela-Arnold & Evans, 2014; Mirman et al., 2008; Saffran, Aslin, et al., 1996; Saffran et al., 1997; Saffran, Newport, et al., 1996; Toro et al., 2005) show that both adults and children perform above chance level in the off-line test phase (but please note that previously found null results may not have been published, as is mentioned by Black and Bergmann (2017). However, often non-words (combinations of syllables that had never occurred as such in the familiarization phase, TP = 0) are used as foils in the test phase. In the current study we used part-words as foils (TP = 0.333), which did occur in the familiarization phase, just less often than the words. Although discrimination between words and part-words has been shown in infant, child and adult studies on word segmentation (Batterink & Paller, 2019; Johnson & Tyler, 2010; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996; Thiessen et al., 2005), in our third experiment we wanted to investigate whether sensitivity to word boundaries would be revealed as a preference for words over non-words (instead of part-words).

2.5 Experiment 3

2.5.1 Methods and materials

Participants. Forty-six adults (35 female, 11 male) between the age of 18;4 and 35;7 (*M* = 22;7, *SD* = 3;5) participated in the study. One participant was excluded because of the use of medicines, leaving 45 participants for data analysis. All remaining participants met the conditions as described in Experiment 1.

Design. The familiarization phase of the word segmentation task was identical to Experiment 2. However, the test phase was changed. Instead of part-words, combinations of syllables that had never occurred in the familiarization phase (non-words, TP = 0) were used as foils. For language A, the non-words *kido*, *moba*, *dati* and *gilu* were constructed and for language B the non-words *bagi*, *timo*, *luda* and *doba*. The foils were constructed by combining the first syllable of a word with the second syllable of another word. Sequences with a double vowel (e.g. *daba*) or that only differed from a target word in one sound (e.g. *kida*) were avoided, and we aimed to construct a set of foils that contained all the syllables from the language. The new foils met the same conditions as the test stimuli in Experiment 1 and 2, and the test phase was constructed the same way.

Procedure. The procedure was identical to that of Experiment 2.

2.5.2 Results

The mean accuracy was 0.54 (*SD* = 0.16). See Figure 2.7 for the accuracy data. A generalized logistic mixed-effects model was conducted (see section 3.2.1). The estimate for the intercept (converted into probability) was 0.55 (95% CI: 0.47 … 0.63), which is not significantly different from chance level: $z = 1.147$, $p = 0.25$. There was a significant effect of Version: the odds that participants who did version B of the task chose the correct answer were 1.85 times (CI: 1.23 … 2.83) higher than for the participants who did version A ($z = 2.974$, $p = 0.0029$). The effect of TargetOrder was not significant. See Table 2.6 for the results of the model. On the basis of these null-results we cannot say whether adults distinguish words from non-words, which would have indicated knowledge of word boundaries, in the off-line test phase of the current word segmentation task.

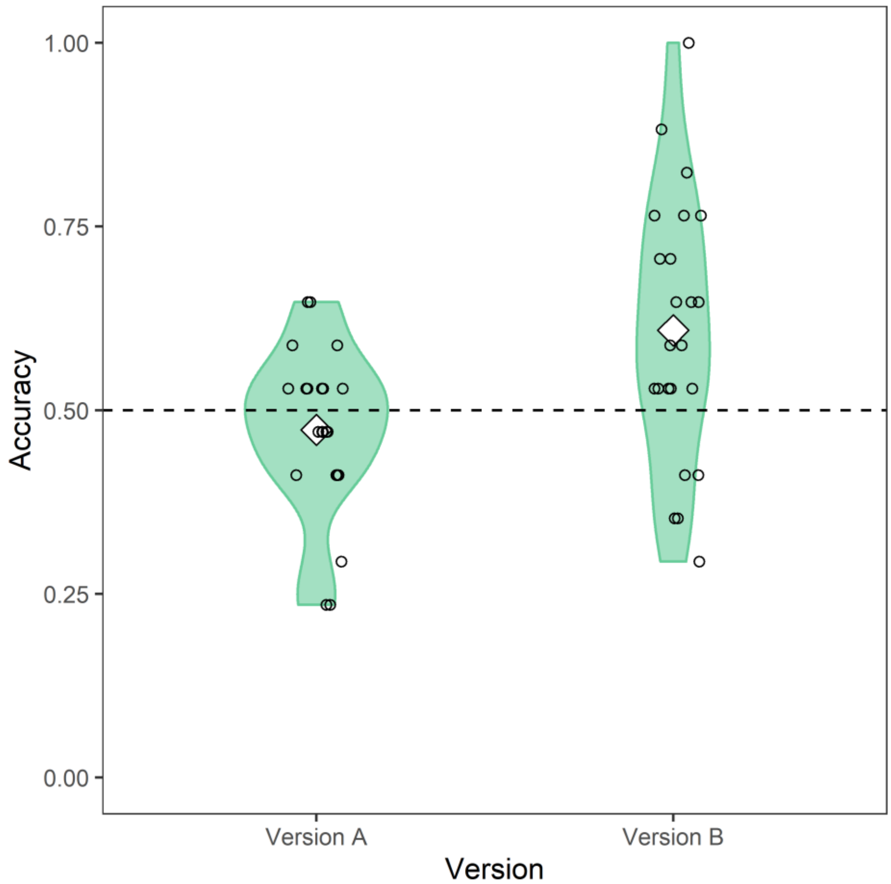


Figure 2.7 – Descriptive plot of participants' accuracy in version A and B of Experiment 3.

Table 2.6 – Results from the linear mixed-effects model.

	Intercept	Version	TargetOrder
Estimate	Probability: 0.55	Odds: 1.85	Odds: 1.66
95% CI	0.47 … 0.63	1.23 … 2.83	0.88 … 3.22
z	1.147	2.974	1.673
p	0.25	0.0029	0.094

2.5.3 Discussion

As in our first two experiments, we did not find evidence that participants were sensitive (or not sensitive) to the statistical regularities in the input of the word segmentation task, even while we may suppose that the discrimination task was easier than in Experiment 1 and 2. There was a significant effect of test version which suggests that word boundaries in language B are somehow easier to learn than those of language A. However, while there was no evidence for lack of balance between the phonotactic probability of the foils and targets in the first two experiments, there was such evidence for Experiment 3. The average phonotactic frequency of the foils of language B (M = 0.769, SD = 0.021) was significantly lower than that of the targets of language B (M = 1.385, SD = 0.036): $t(6)$ = 5.172, p = 0.0021, while there was no significant difference between the foils of language A (M = 1.179, SD = 0.341) and the targets of language A (M = 1.425, SD = 0.174): $t(6)$ = 1.285, p = 0.246). Thus, for language B it could be easier to choose targets over foils, because the targets sound more “Dutch-like” than the foils. It might be the case that the higher performance in version B of Experiment 3 does not (only) reflect learning of the word boundaries, but a bias that is inherent to the stimuli of the test phase.

2.6 General discussion

In the present set of studies, we aimed to find a method for measuring word segmentation on-line that was suitable for testing school-aged children, but we encountered some unexpected outcomes. Firstly, the click detection task does not seem to be a reliable measure of on-line word segmentation. There was a large amount of individual variation in the data, and test version may have influenced the results. We cannot state that responding to click sounds in a stream of interrupted syllables consistently revealed an effect of word boundary knowledge. The fact that our artificial language consisted of bisyllabic words means that the click sounds occurring within words always divided the words in two single syllables. This is not the case for a language containing trisyllabic words (Franco et al., 2015; Gómez et al., 2011), as within-word clicks in that case still leave two syllables of that word uninterrupted. This repetition of high transitional probabilities (between syllable 1 and 2 and between syllable 2 and 3) might counter the effect of clicks being perceived as cues for word boundaries. Future studies might adapt our paradigm (adding blocks with only a few click sounds to the familiarization phase) using trisyllabic words to test whether the click detection task would reveal on-line learning then. A general difficulty with the click detection paradigm might be the so-called "auditory streaming effect" or "auditory stream segregation" (Micheyl et al., 2010; van Noorden, 1975): participants might perceive the syllables and the click sounds as two separate sound streams. If this is the case (for some listeners), this might be the reason that the position of the click sounds does not influence their reaction time to them. Participants reported that they found it very hard to pinpoint in which specific syllable a click occurred. Moreover, the addition of click sounds to the stream could have distracted the listeners' attention away from the to-be-learned word boundaries, as suggested by Franco et al. (2015) and Toro et al. (2005). The below-chance performance on the off-line test phase in our first experiment seemed to point in that direction. We wanted to investigate this by conducting Experiment 2, in which we tested the same word segmentation task without the addition of the click detection task.

Contrary to our expectations, we did not find evidence for or against sensitivity to words boundaries in adults or children in Experiment 2. This was also the case for our third experiment, in which the foils in the test phase were not part-words (TP = 0.333) but non-words (TP = 0). There are multiple factors that could have influenced our results. First, it could be the case that two-syllable words are somehow too short to "trigger" a statistical learning mechanism, although Graf Estes and Lew-Williams (2015) and Haebig et al. (2017), who also used bisyllabic words, did find a learning effect. However, differently from Haebig et al. (2017), our participants listened to the language for 8 minutes instead of 4.75 minutes. It is possible that the high amount of exposure to the syllables had given the participants the impression that the language consisted of monosyllabic words instead of bisyllabic words, resulting in less sensitivity to syllable combinations. Second, we used natural modified speech instead of synthesized speech. In a meta-analysis, Black and Bergmann (2017) found that infants' word segmentation ability was stronger in experiments that use synthesized speech. Natural speech contains more information than synthesized speech which could make the processing of the stream and consequently learning of word boundaries more difficult. Regarding the test phase, almost all participants stated that they found it difficult and often also reported that the test became more difficult as it progressed. This might be due to the repetition of the targets and foils in the test phase, which could overwrite the (weak) representations that might have been built during the familiarization phase (Siegelman, Bogaerts, Kronenfeld, et al., 2018).

Influence of prior linguistic knowledge could also have played a role (Finn & Hudson Kam, 2008; Siegelman, Bogaerts, Elazar, et al., 2018; van Hedger et al., 2022). Participants, especially adults or older children, who are subjected to an artificial language in a word segmentation task are not blank slates but already have linguistic knowledge and thus expectations about sounds and sound combinations. This knowledge might influence the learning process. As this influence is hard to predict correctly, the particular words that are chosen in an experiment might impact participants' performance. This is for example illustrated by findings of Erickson et al. (2016), who tested participants on two word segmentation tasks with different sets of words. Performance on one task did not reliably predict performance on the other task. Siegelman et al. (2018) suggest that the influence of prior linguistic knowledge (or "entrenchment effects") plays a very important factor in the large differences in effect sizes and reliability that is found between statistical learning studies. The influence of prior linguistic knowledge might in some cases be stronger than the influence of the statistical properties of the input that participants are briefly subjected to. Entrenchment effects could have led to the null results and the unexpected version effects in Experiment 1 and 3 of the current study and possibly more studies that have ended up in drawers. Future research should investigate this phenomenon in depth.

In sum, measuring word segmentation ability reliably might be more sensitive to methodological choices than assumed (see also Black and Bergmann, 2017). We would like to emphasize that studies that fail to find a significant effect may not be published (the "file drawer effect"). Access to null results is essential for reliable meta-analyses, which are an important source of empirical evidence. Therefore, it is important to report the results of our current study. Future research should systematically investigate what constraints the word segmentation ability and how (on-line) learning can be detected reliably (Siegelman et al., 2017). For example, Lukács et al. (2021) and Lukics and Lukács (2021) tested out a method of measuring word segmentation on-line with a syllable detection task, and Kidd et al. (2020) used a serial recall task to measure off-line learning more reliably.

Chapter 3

Implicit cross-situational word learning in children with and without developmental language disorder and its relation to lexical-semantic knowledge

This chapter is a slightly modified version of the submitted article:

Broedelet, I., Boersma, P., & Rispens, J. (2022). Implicit cross-situational word learning in children with and without developmental language disorder and its relation to lexical-semantic knowledge. *Submitted to Frontiers in Communication: Language Sciences.*

R scripts, data and materials are available on FigShare:
https://doi.org/10.21942/uva.c.6152406

Abstract

Research indicates that statistical learning plays a role in word learning by enabling the learner to track the co-occurrences between words and their visual referents, a process that is named cross-situational word learning. Word learning is difficult for children with developmental language disorder (DLD). A deficit in statistical learning has been suggested to contribute to the language difficulties in these children. In the current study we investigate whether children with DLD have more difficulty than TD children with learning novel word–referent pairs based on cross-situational statistics in an implicit task, and whether this ability is related to their lexical-semantic skills. Moreover, we look at the role of variability of the learning environment. In our implicit cross-situational word learning task, each trial in the exposure phase was in itself ambiguous: two pictures of unknown objects were shown at the same time and two novel words were played consecutively, without indicating which word referred to which object. However, as every word occurred with its correct referent

consistently, the children could learn the word–referent pairs across trials. The children were not explicitly instructed to learn the names of new objects. As an on-line measure of learning, eye-movements were recorded during the exposure phase. After exposure, word–referent knowledge was also tested using multiple choice questions. Different measures of lexical-semantic knowledge were administered to the children with DLD, as well as tasks measuring non-verbal intelligence and phonological processing. Contextual variability (the number of different distractors with which a particular word–referent pair occurs across trials) was manipulated between subjects by constructing two types of exposure conditions: low contextual diversity vs. high contextual diversity. Both groups of children performed significantly above chance level on the test phase, but the TD children significantly outperformed the children with DLD. This indicates that children with DLD have more difficulty with implicit cross-situational word learning. We found no significant effect of contextual diversity. The eye-tracking data revealed evidence of on-line learning, but no differences between groups. The regression analyses did not reveal any significant predictors of off-line or on-line cross-situational word learning ability.

3.1 Introduction

Young children learn a large number of words in a relatively short period of time, knowing an estimate of 14.000 words at six years old (Suanda et al., 2014). How are they able to do this? This question is especially puzzling considering the referential ambiguity problem that children often experience (Quine, 1960): children hear a word unknown to them and see multiple potential referents at the same time. How do they learn to match words to their correct referents? Recent research into cross-situational word learning has indicated that statistical learning plays a role in tracking co-occurrences between words and their corresponding referents. This type of learning may thus be important for word learning (Kachergis et al., 2014; L. Smith & Yu, 2008; Suanda et al., 2014; Yu & Smith, 2007; Yurovsky et al., 2014).

Learning words requires more effort for some children than for others. Children with developmental language disorder (DLD) often have difficulties with the development of word knowledge (Brackenbury & Pye, 2005; McGregor et al., 2013; Nation, 2014; Sheng & McGregor, 2010). Evidence suggests that an impairment in statistical learning, a learning mechanism that supports the extraction of patterns and regularities from sensory input, contributes to the language difficulties in these children (Evans et al., 2009; Haebig et al., 2017; Hedenius et al., 2011; Hsu & Bishop, 2014; Lammertink et al., 2017; Mainela-Arnold & Evans, 2014). The current study aims to investigate whether a cross-situational word learning deficit might (partially) explain their hampering lexical acquisition. Difficulty in tracking the co-occurrences between words and their corresponding referents might result in a problematic vocabulary development in children with DLD.

On the basis of accuracy and eye-tracking data, we investigate whether children with DLD have more difficulty than typically developing (TD) children when learning word–referent pairs in a cross-situational word learning experiment, as well as whether this cross-situational word learning ability is related to different types of vocabulary knowledge in children with DLD. Moreover, as previous research has shown that high variability in the learning environment might enhance statistical learning (Grunow et al., 2006; von Koss Torkildsen et al., 2013), we manipulate the contextual diversity of the to-be-learned word–referent pairs in our experiment to investigate whether this affects cross-situational word learning in children with and without DLD.

3.2 Background

3.2.1 Lexical acquisition and cross-situational word learning

The acquisition of vocabulary starts in early infancy, continuing throughout life. The acquisition of a rich vocabulary entails several competences, including (but not limited to) the discovery of word forms, learning about concepts and word meanings, word–meaning association and the expansion of lexical representations (Ralli et al., 2010; Yu & Ballard, 2007). Previous studies have shown that statistical learning likely contributes to at least part of these processes. For example, segmenting words from running speech in which word boundaries are not consistently indicated is supported by statistical learning mechanisms (Graf Estes, 2009; Saffran, Aslin, et al., 1996; Saffran et al., 1997). In addition to defining a word, children also need to map those word forms to their corresponding referents in the real world. The ability of fast mapping is important in early word learning (see Horst and Samuelson (2008) for a review). Fast mapping is described as an "all-or-none" learning mechanism for which one single exposure to a new word and its corresponding referent is sufficient to link them in memory. However, this is just the start of building elaborate lexical entries. After fast mapping, the meaning and scope of the word needs to be further specified and the words needs to be placed in a broader network of related words, a process also called "slow mapping" (Blythe et al., 2010; Carey, 1978), which requires repeated exposure to words. Moreover, the research by Horst and Samuelson (2008) suggests that fast mapping is not sufficient for long-term word knowledge: after a mere five minutes, the 24-month-old infants in their fast-mapping experiment could no longer express what they had learned.

Word-learning contexts outside the lab are usually much more ambiguous than in controlled experiments. The fast speech stream and the visual world that young children encounter contain many (new) words and many different potential referents. How does a child learn the correct word–referent mappings? Referential ambiguity, or the word-to-world mapping problem, has been described by Quine (1960) and many others: "In any naming event, a novel word can refer to any object present, its properties, the speaker's feelings or intentions for it, an impending action, or something else altogether" (McMurray et al., 2012, p. 832). Thus, real-life learning situations might often not be ideal fast-mapping situations.

Another factor that needs consideration is that the number of possible referents is constrained by built-in or learned biases. For example, words usually refer to whole objects rather than parts or properties of an object (MacNamara, 1972), and children know that a novel word should not be linked to a referent that already is linked to another word (mutual exclusivity; Markman & Wachtel, 1988). Moreover, identifying the attentional focus of the speaker (Baldwin, 1991) and syntactic bootstrapping (Gleitman et al., 2005) reduces referential ambiguity (Blythe et al., 2010). Although these biases play an important role in word learning, evidence suggests that statistical learning mechanisms could be used on top of that to exploit an environment in which there is often a degree of (referential) uncertainty. Word–referent mapping could be viewed as a gradual, accumulative process: the learner can reduce referential uncertainty and extend meaning representations when a word is encountered in different contexts. This would mean that children can make use of ambiguous learning situations rather than only learn when there is no ambiguity at all (Yurovsky et al., 2014). Thus, words are not (always) learned in one single, unambiguous event. Rather, children use statistical learning, that is the ability to use information about the co-occurrence of words and referents from many different encounters, to acquire a vocabulary network. See Figure 3.1 for a visual representation of cross-situational word learning.

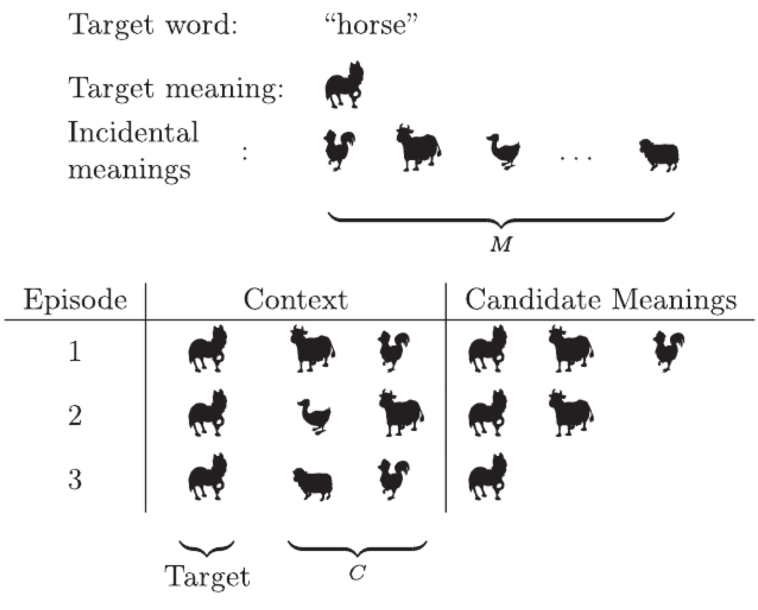


Figure 3.1 – Visual representation of cross-situational word learning. Picture from Blythe et al., 2010.

In cross-situational word-learning tasks (Smith & Yu, 2008; Yu & Smith, 2007), participants are usually exposed to multiple novel words and multiple novel objects per learning trial (see *Figure 3.2*). In these tasks, each individual learning trial is *in itself* ambiguous, as multiple words and referents appear simultaneously, with no indication as to which word should be mapped to which referent. However, as the correct word–referent pairs do consistently occur together, the correct mappings can be learned by accumulating evidence *across trials*. This task is a (strongly) simplified simulation of real word-learning situations in which there is often some amount of referential ambiguity. Studies have shown that adults (Fitneva & Christiansen, 2011; Kachergis et al., 2014; Smith et al., 2011; Suanda & Namy, 2012; Yu & Smith, 2007), infants (Smith & Yu, 2008; Vlach & Johnson, 2013; Yu & Smith, 2011)and 5–7-year-old children (Suanda et al., 2014; Vlach & DeBrock, 2017) are able to learn word–referent pairs in this paradigm, after only a few minutes of exposure. Eye-tracking has been used as a measure of on-line learning in cross-situational word-learning tasks in adults (Fitneva & Christiansen, 2011; Yu et al., 2012), infants (Yu & Smith, 2011), children with and without autism (Venker, 2019) and children with and without DLD (Ahufinger et al., 2021), allowing for more fine-grained analyses of learning. Besides behavioural experiments, computational models show that cross-situational learning mechanisms could explain learning large vocabularies in a relatively short amount of time in spite of referential ambiguity (Blythe et al., 2010; Yu & Smith, 2012), although it is still unclear whether associative learning, hypothesis testing strategies or both may underlie this ability.

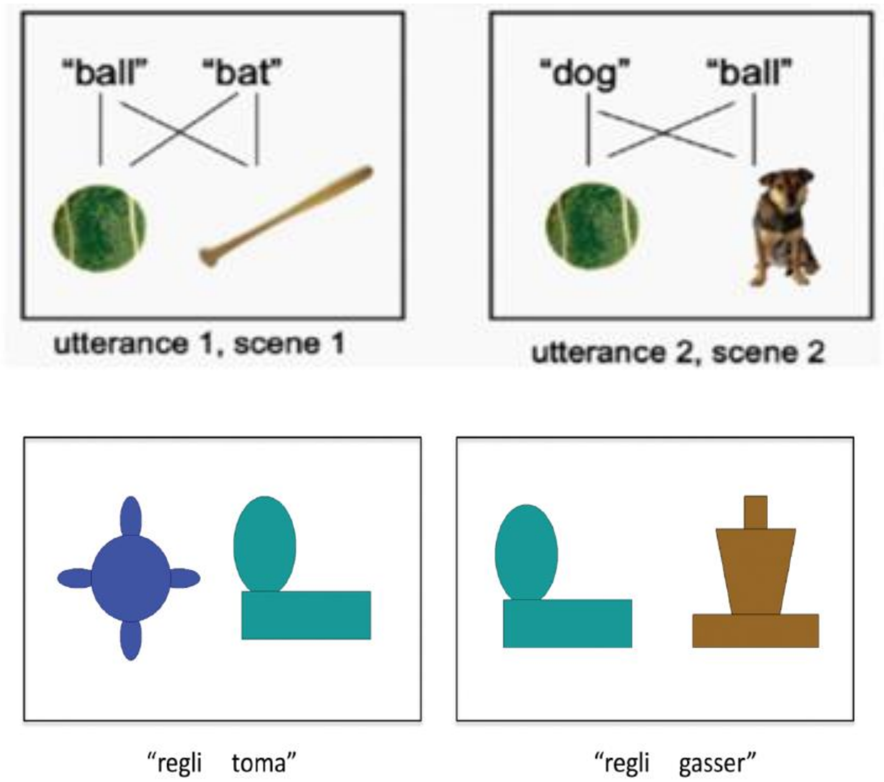


Figure 3.2 – Cross-situational learning in the experiment of Smith and Yu (2008).

3.2.2 Lexical deficits in children with DLD

Children with DLD have evident problems in the development of the lexicon, although grammatical problems are generally more apparent (Brackenbury & Pye, 2005; Jackson, Leitão, et al., 2019; Nation, 2014). These reported vocabulary problems may last into adulthood (McGregor et al., 2017) and include smaller vocabulary size and more superficial word knowledge (McGregor et al., 2013), less accurate word naming (Dockrell et al., 2001; Lahey & Edwards, 1999; Leonard et al., 1983; McGregor, 1997; McGregor et al., 2002), impoverished semantic representations (Dockrell et al., 2003; Drljan & Vuković, 2019; Mainela-Arnold et al., 2010; Marinellie & Johnson, 2002) and less efficient lexical-semantic networks (Drljan & Vuković, 2019; McGregor et al., 2012; Sandgren et al., 2021; Sheng & McGregor, 2010).

Learning new words is generally problematic in this population (Alt & Plante, 2006; Haebig et al., 2017; Kan & Windsor, 2010; Kapa & Erikson, 2020; Nash & Donaldson, 2005). For example, Alt and Plante (2006) reported that children with DLD have difficulty with learning phonological forms of words as well as learning about semantic properties of words such as colour and shape. Evidence also suggests that fast mapping is difficult for children with DLD (Alt et al., 2004; Gray, 2004; Haebig et al., 2017; Jackson, Leitao, et al., 2019; Rice et al., 1994; see Jackson, Leitão, et al. (2019) for a review of the different types of word learning experiments that have been tested in individuals with DLD).

Thus, the lexical-semantic development in children with DLD is affected, but the underlying cause of these difficulties is still under debate. Phonological short-term memory has often been reported as being more limited in children with DLD compared to TD children, and has been hypothesized to contribute to their lexical-semantic deficits (see Montgomery et al., 2010 for a review). Indeed, phonological short-term memory has been shown to impact fast mapping in children with DLD (Alt & Plante, 2006; Jackson, Leitao, et al., 2019), and results by Quam et al. (2020) indicate that sound discrimination ability affects word–object mapping in children with DLD. Whether there is actually a causal relation between phonological memory and lexical abilities is still debated (Melby-Lervåg et al., 2012).

3.2.3 Statistical learning deficit in DLD

Another strand of DLD research focuses on learning mechanisms that are not specific to language. A growing body of evidence implies that impaired statistical learning underlies the language deficits in children with DLD. These children seem to have more difficulty with extracting patterns from their environment, for example when extracting words from running speech in a word segmentation task (Evans et al., 2009; Haebig et al., 2017), learning non-adjacent dependencies in an artificial grammar learning task (Hsu et al., 2014; Lammertink et al., 2019), or learning motor sequences in a serial reaction time task (Lukács & Kemény, 2014). Meta-analyses also point into the direction of a statistical learning deficit in children with DLD (Lammertink et al., 2017; Lum et al., 2014; Obeid et al., 2016). Please note, however, that some studies report no evidence for (or against) a statistical learning deficit (Aguilar & Plante, 2014; Lammertink, Boersma, Rispens, et al., 2020; Lammertink, Boersma, Wijnen, et al., 2020; Noonan, 2018). Importantly, statistical learning ability has been shown to be correlated with language abilities in TD children (Conway et al., 2010; Ellis et al., 2014; Hamrick et al., 2018b; Kaufman et al., 2010; Kidd, 2012; Kidd & Arciuli, 2016; Misyak et al., 2010; Newman et al., 2006; Shafto et al., 2012; Spencer et al., 2015) and in children with DLD (Evans et al., 2009; Hedenius et al., 2011; Mainela-Arnold & Evans, 2014; Misyak et al., 2010; Tomblin et al., 2007).

As discussed above, statistical learning seems to play an important role in the development of the lexicon. This hypothesized relation is underlined by findings of positive correlations between statistical learning ability and vocabulary size (Spencer et al., 2015) and even more strongly by the finding of predictive relationships between statistical learning and later vocabulary development in longitudinal studies (Ellis et al., 2014; Shafto et al., 2012; Singh et al., 2012). In children with DLD, this relationship has also been found (Evans et al., 2009; Mainela-Arnold & Evans, 2014). However, the relationship between statistical learning and specifically the largely unexplained lexical-semantic difficulties in children with DLD is not yet clear. The cross-situational word-learning paradigm offers a way to investigate the role of statistical learning in finding a word's meaning.

Cross-situational word learning has only sparsely been investigated in children with DLD. However, incidental word learning has been studied using the quick incidental word learning (QUIL) paradigm, which aims to mimic naturalistic word learning (Rice et al., 1990). In these tasks, new words are not explicitly taught but embedded in video stories. Children with DLD learn fewer words in such tasks (see Chung and Yim, 2020 for a summary). Findings by Rice et al. (1994) indicate that children with DLD are able to learn new word–referent mappings in a QUIL task, but need more exposure to the words than TD children. Correlations between QUIL ability and language skills have been reported by Gordon et al.

(1992) and Yang et al. (2013). In a recent study, Chung and Yim (2020) investigated QUIL in 4–6-year-old children with and without DLD, also measuring eye movements during learning. In the task, the children were exposed to a 5-minute-long video story in which five novel words had been embedded in sentences, each word three times, without further instructions. Afterwards it was tested whether the children could pick the right object corresponding to the novel words. Results showed that children with DLD score lower on this task, suggesting that they learn fewer words from watching this video. Moreover, the eye-tracking data revealed that children with DLD fixate less often on these target objects over time, while the fixations of TD children increase over time, and their looks are more widely scattered in general. As fixation time predicts word learning, the gazing pattern of children with DLD seems to reflect their difficulty linking new words to their referents.

The study by Ahufinger et al. (2021) was the first to directly test children with DLD on the ability of tracking the co-occurrences between multiple words and visual referents in a cross-situational word learning task. In their experiment, children with and without DLD were subjected to a familiarization phase in which they could learn the names for eight robot-like figures in 16 trials. In each learning trial, the participants saw two pictures and heard two words, without any indication as to which word referred to which picture. After familiarization, the children were tested twice on each word–referent pair using four-alternative forced choice questions. Moreover, eye movements were measured during the familiarization phase and the testing phase. Although both groups of children performed significantly above chance level on the testing phase, the children with DLD had learned significantly fewer word–referent pairs than the TD children. The eye-tracking data did not reveal any preferences for target or distractor items during the familiarization phase, nor significant group differences in looking behaviour. Eye-movements during the test phase were interpreted by the authors as a measure of the confidence children had in their answer. When only trials in which the participant had given the right answer were included in the analysis, the TD children looked significantly longer towards the target image than the children with DLD, indicating that TD children are more confident in their answer than children with DLD.

McGregor et al. (2022) investigated cross-situational word learning in children with DLD, with the aim of examining how learner characteristics influence this ability. In their cross-situational word learning task, overt responses were recorded during the learning phase of the experiment, enabling the researchers to track learning during the exposure to word–referent pairs. Accuracy on word–form retention (recognizing a trained word from three possibilities, for example *zote, zoke or zofe*) and word–referent retention (matching the correct picture to a trained word) was also measured after a 5-minute interval. In every learning trial, the children saw two pictures and heard one word, and were prompted to choose the correct picture. Results show that children with DLD are less accurate at picking the correct referent during the learning phase compared to TD children. A significant main effect of Trial indicates that children get better at picking the right picture during the learning phase. However, there was no evidence for a slower learning trajectory for children with DLD, as the interaction between Trial and Group was not significant. In the test phase, the TD children also significantly outperformed the children with DLD. This was the case for both word form recognition and word–referent link recognition, and there was no evidence for a difference in performance on those two tasks. Moreover, links between vocabulary, attention, phonological working memory and cross-situational word learning were investigated. Vocabulary was the strongest predictor of cross-situational word learning. A relative importance analysis, a type of analysis that can be used to determine unique variance in a dependent variable and is suitable when predictors are correlated, indicates that this link is stronger for TD children than for children with DLD. As sustained attention was a significant predictor of the children with DLD's performance in the final learning block, this ability appears to contribute to cross-situational word learning in children with DLD, but please note that sustained attention in itself was not a significant predictor of performance in the relative importance analysis.

The studies of Ahufinger et al. (2021) and McGregor et al. (2022) show that children with DLD indeed have more difficulty with statistical word–referent mapping. As the task of Ahufinger et al. (2021) included an extensive explanation and practice phase before familiarization, we do not yet know how children with DLD perform compared to TD children on *implicit* cross-situational word learning (see Evans et al., 2009 for a study on implicit word segmentation in children with and without DLD). As Ahufinger et al. state themselves, "(…), these explicit instructions may have triggered a compensatory mechanism (Ullman & Pullman, 2015) to help children with DLD to perform above chance. This hypothesis, however, should be further investigated by assessing the accuracy in this population in a CSSL task with no explicit instructions and no explicit response." (p. 14). Moreover, our study addresses the relationship between cross-situational word learning ability and different measures of lexical-semantic knowledge in children with DLD. McGregor et al. (2022) implemented a behavioural measure of on-line learning. As we aim to investigate implicit cross-situational word learning, measuring eye-tracking is more suitable for our study.

Variability in the learning environment seems to enhance statistical learning: previous research shows that people often learn better on tasks tapping statistical learning when variability is increased in some way. For example, Gómez (2002) tested artificial grammar learning in adults and 18-month-old infants. The grammar consisted of non-adjacent dependencies (for example: *pel* X *jic*). Participants were significantly better at learning the dependency relation between *pel* and *jic* when the intervening element (X) had 24 unique forms, than when X had only 12 different forms. Using the same task, Grunow et al. (2006) found that high variability of the X element also seems to have a positive effect in adults with and without language-based abilities. Other studies also indicate that variability has a positive effect on learning on both individuals with and without language-based disorders, and that more variability results in better generalization of the learned information (Aguilar et al., 2017; Desmottes et al., 2017; Perry et al., 2010; Plante et al., 2014; von Koss Torkildsen et al., 2013). Variability in the learning environment (but not in the to-be-learned target itself) might cause the invariable target or pattern to stand out more and therefore it becomes easier to learn.

Increasing variability in the learning context has also been applied to cross-situational learning tasks, by manipulating the contextual diversity of the to-be-learned word–object pairs. Contextual diversity in this case is defined as "the number of different sets of stimuli with which each word–object pairing co-occurs across learning trials" (Suanda et al., 2014, p. 397). Suanda and Namy (2012) found that greater contextual diversity enhances the learning of word–object mappings in adults: items that occur in more variable contexts (with more different distractor items) are easier to learn than items in a less variable context. Similarly, Suanda et al. (2014) made a comparison between high, moderate and low contextual diversity conditions in a cross-situational learning experiment, and found that contextual diversity enhances cross-situational learning in children of 5-7 years old. In the current research, it is tested whether contextual variability enhances cross-situational word learning differently in TD children than in children with DLD.

3.3 The current study

Our study aims to shed light on the relationship between cross-situational word learning and lexical-semantic knowledge in children with and without DLD. To this end, we investigate implicit cross-situational word learning in 7-9-year-old children with and without DLD, as well as the relation between this ability and various lexical-semantic skills in children with DLD. To investigate cross-situational word learning, we use off-line as well as on-line measures. The children's eye movements are measured during the familiarization phase to gain insight of how learning of word–referent pairs unfolds. There has been the need for on-line measures of statistical learning, because off-line measures such as performance on a testing phase are not always a reliable measure of statistical learning ability (Siegelman, Bogaerts, Kronenfeld, et al., 2018). However, to the best of our knowledge, previous studies have not looked at the development of looking times towards the target image *across trials*. We consider this measure as reflecting learning of word–referent pairs during the experiment. We aim to answer the following research questions:

- RQ1A: do children with DLD have more difficulty than TD children with learning word–referent pairs in an implicit cross-situational word learning task?
- RQ1B: do children with DLD show weaker on-line learning than TD children during implicit cross-situational word learning?
- RQ2: is cross-situational word learning ability related to lexical-semantic skills in children with DLD?
- RQ3A: does higher contextual diversity enhance cross-situational word learning?
- RQ3B: does contextual diversity impact cross-situational word learning differently in TD children than in children with DLD?[6]

We expect to find that children with DLD are less proficient in cross-situational word learning than TD children, which will be reflected by both behavioural and eye-tracking data. Moreover, we expect that cross-situational word learning ability is related to lexical-semantic knowledge in children with DLD. Finally, we expect that contextual diversity enhances learning in both groups of children. We have no hypothesis about a group difference on this enhancing effect of contextual diversity and thus explore whether this is the case.

3.4 Method

The task of the current study, based on Smith and Yu (2008) and Suanda et al. (2014), amongst others, is designed to measure cross-situational word learning in school-aged children (7–9 years old). Learning is tested off-line

[6] Since we posit multiple research questions, we adjust the significance criterion to $p = 0.01$ as opposed to the conventional $p = 0.05$.

(test phase after familiarization) and on-line (eye-tracking during familiarization). Moreover, the influence of contextual diversity on word learning is investigated.

3.4.1 Participants

Twenty-six children diagnosed with DLD (18 boys and 8 girls) between the age of 7;2 (years;months) and 9;3 were tested (average: 8;1). As a control group, we used previously collected data of 26 TD children (15 boys and 11 girls) between 7;6 and 8;11 (average: 8;2)[7]. The subgroup was selected from a larger sample to match with the DLD group regarding age, gender and the condition of the experiment (contextual diversity). All children had normal or to-normal-corrected vision, and did not have hearing loss or a diagnosis of AD(H)D or ASD. At least one of the caretakers had acquired Dutch as a native language. The TD group did not have any history of language disorders or dyslexia. The Ethical Committee of the Faculty of Humanities of the University of Amsterdam approved the experiment. Caretakers of the children gave written informed consent for their participation.

All children in the DLD group had been previously diagnosed with DLD by a professional speech and language therapist and met the inclusion and exclusion criteria used within the institution from which they were recruited (Pento, Royal Dutch Auris Group and VierTaal). Using data collected by the institutions, it was checked that all children scored at least 1.5 standard deviations below the age norm on at least two out of four language domains (speech, auditory processing, grammar,

[7]We had planned to test an age-matched group of TD children. Unfortunately, we were unable to administer the tests as all primary schools in the Netherlands were closed from March to June 2020 due to the outbreak of COVID-19. After the reopening of the schools many restrictions still applied, making it impossible to enter schools for testing participants. We therefore decided to use a subset of an already collected pilot data as control data. No articles based on this data have been published. As a result of this, the control group, unlike the DLD group, was not tested on the background tasks measuring vocabulary, morphosyntactic skills, phonological processing and non-verbal intelligence. This means the control group could unfortunately not be matched on vocabulary skills to the DLD group.

vocabulary), measured using standardized tests. Furthermore, their language problems were not secondary to neurological of physiological disorders such as ASD, ADHD, a severe form of dyspraxia, hearing difficulties or genetic syndromes like Down syndrome or 22q11 syndrome.

3.4.2 Stimuli

Eight novel words and eight novel objects were used to form word–referent pairs. The novel objects were taken from the database of Kachergis et al. (2014), with permission from the authors. All objects were uncommon, difficult to name objects (see Figure 3.3).

Figure 3.3 – Novel objects used in the experiment (from Kachergis et al., 2014).

The novel words (/dita/, /loχa/, /mɪp/, /kasi/, /ʋɛfəl/, /sulɛp/, /reχɛs/ and /χɔp/) were based on the novel words conducted by Lammertink et al. (2019). All words sounded like Dutch words and were recorded in a neutral manner by a female native speaker of Dutch.

3.4.3 Design

The familiarization phase consisted of 28 trials, in which eight word–referent pairs could be learned. A word co-occurred with its referent on seven trials in total. In each trial two word–referent pairs were presented (see Figure 3.4). Each trial in itself was ambiguous in the sense that it was not indicated which of the two words referred to which of the two referents. The position of the objects (left/right) and the order of the words (said first/second) was varied: in half of the trials the first word corresponded to the left object and the second word to the right object ("congruent" trials, named so because the reading direction in Dutch is from left to right), while in the other half of the trials the first word corresponded to the right object and the second word to the left object ("incongruent" trials). Each word consistently appeared with its corresponding referent.

For every participant, words were paired with objects randomly. Thus, for one participant /dita/ could refer to the spiral-like object, while for another it could refer to the white round object. The order of the learning trials was pseudo-randomized such that an object could not occur on the same side of the screen two times in a row and a word could not occur as the first/second word two times in a row. Trials lasted five seconds in total, resulting in a familiarization phase of approximately 2 minutes and 20 seconds, but please notice that the exact duration of the learning phase varied between participants, as they could only proceed to the next learning trial if they were looking towards the screen. In every trial, the two objects appeared on the screen two seconds before the first word played. All words had a duration of one second, and a one-second silence was placed between the two words. The trial structure is similar to that used by Smith and Yu (2008), but the time before the onset of the first word and the time between words was extended, so that participants had more time to process the words and the objects. Eye movements were measured during the familiarization phase to measure on-line learning.

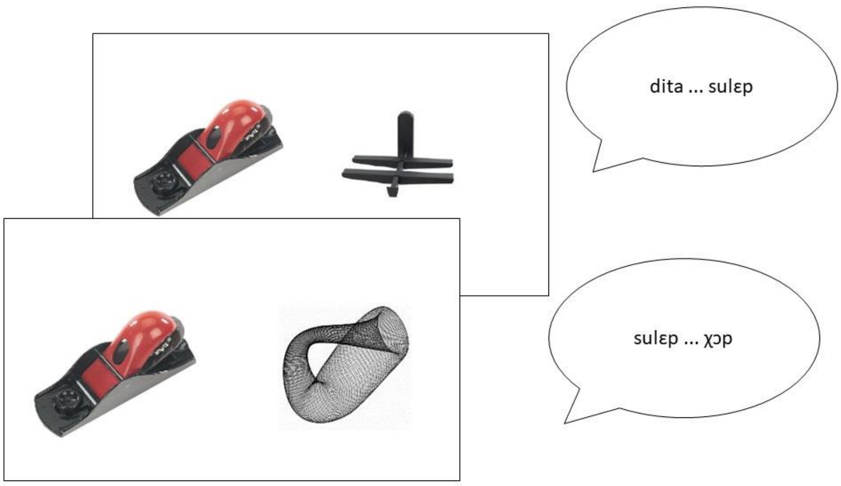


Figure 3.4 – A familiarization trial.

Contextual diversity was manipulated between participants. In both conditions, eight word–referent pairs could be learned across trials, but the conditions differed in the variability of the environment in which the word–referent pairs occurred. In each learning trial, two word–referent pairs were presented simultaneously. In the high contextual diversity condition (*high-CD*), a particular word–referent pair (for example word 1 and picture 1, pair 1−1) occurred with a different word–referent pair each time across trials (with pairs 2–2, 3–3, 4–4, 5–5, 6–6, 7–7 and 8–8). However, in the low contextual diversity condition (*low-CD*), the accompanying word–referent pairs were sometimes the same. For example, pair 1−1 occurred with pair 2–2 three times, with pair 3–3 three times and with 4–4 once. Thus, in the low-CD condition there was less diversity across trials. See Table 3.1 for the combinations of word–referent pairs in the two familiarization conditions.

In the test phase, all eight word–referent pairs were tested once. Participants heard a word three times and had to choose between four objects which was the correct one. The same audio files as in the familiarization phase were used. In the high-CD condition, three random objects are chosen as foils. All these foils had occurred with the word equally often (once). In the low-CD condition, the three foils that had occurred with the word are chosen as foils. Two of the foils had occurred with the word three times, and one foil had occurred with the word once (see Table 3.1).

Table 3.1 – High-CD and low-CD familiarization conditions. In the high-CD condition, word 1 was presented with its referent (object 1) seven times. In these seven trials, all other objects occurred once. In the low-CD condition, only objects 2, 3 and 4 occurred with this word–referent pair.

High contextual diversity									Low contextual diversity								
word→ ↓object	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	word→ ↓object	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	7	1	1	1	1	1	1	1	**1**	7	3	3	1				
2	1	7	1	1	1	1	1	1	**2**	3	7	3			1		
3	1	1	7	1	1	1	1	1	**3**	3	3	7				1	
4	1	1	1	7	1	1	1	1	**4**	1			7	3			3
5	1	1	1	1	7	1	1	1	**5**				3	7	3		1
6	1	1	1	1	1	7	1	1	**6**		1			3	7	3	
7	1	1	1	1	1	1	7	1	**7**			1			3	7	3
8	1	1	1	1	1	1	1	7	**8**				3	1		3	7

3.4.4 Apparatus

The experiment ran in E-Prime 3.0 (Psychology Software Tools Inc, 2016) on a Windows laptop computer with a 17-inch monitor. Eye movements of the participants were measured with a Tobii Pro X2-120 mobile eye-tracker which was attached to a laptop. Gaze data were recorded at 120 Hz (120 samples per second).

3.4.5 Background measures

The cross-situational word learning task was part of a larger test battery. A number of background measures that tap into different types of linguistic skills and other cognitive skills were administered to the children with DLD.

Language measures. In the language domain we tested different types of lexical skills and morphosyntactic skills using subtests of the CELF-4-NL (Clinical Evaluation of Language Fundamentals: Core Language Scales, Dutch version; Semel et al., 2010) and the Peabody Picture Vocabulary Task-III-NL (PPVT; Schlichting, 2005). Regarding the lexical skills, we used the Expressive Vocabulary task to measure expressive vocabulary, the Word Classes task (part 1 or 2, depending on the age of the participant) to measure the ability to express relationships between words, the Word Associations task to measure the ability of recalling words in a certain semantic category (all three subtasks of the CELF-4-NL), and the PPVT to measure passive vocabulary size. Moreover, morphosyntactic knowledge was measured using the Sentence Recall subtest from the CELF-4-NL, and the non-word repetition task (Rispens & Baker, 2012) was administered to test verbal short-term memory.

Cognitive measures. We administered the Raven Progressive Matrices (Raven et al., 2003) to measure non-verbal intelligence. Auditory short-term memory (Number Repetition 1 from the CELF-4-NL, digit span task) and working memory (Number Repetition 2, from the CELF-4-NL, digit span backwards task) were administered as well.

3.4.6 Procedure

For every participant, the experiment consisted of a calibration phase, a familiarization phase and a test phase. Participants sat behind the computer screen. The calibration procedure was run with E-Prime. As a first part of the calibration, it was checked whether participants' gaze was in the centre of the screen, and if necessary the position of the laptop was adjusted. The calibration procedure included nine fixation points and took approximately 2 minutes. After calibration, the task was explained to the participants. A cute alien was shown on the screen and pre-recorded child-directed instructions were played. Participants were instructed to look carefully at the screen and listen carefully to the words, and they were told that there would be some questions at the end of the experiment. Thus, it was not explicitly explained that they should learn word–referent pairs.

Participants were then exposed to either the high-CD or the low-CD familiarization condition. Every learning trial started with a fixation cross (a + in the middle of the screen). Participants automatically proceeded to the learning trial if they looked at the cross for 200 consecutive milliseconds (24 samples). A cover task was added to the familiarization phase to make sure the participants kept paying attention.

The same alien that gave them instructions appeared jumping on the screen at five random moments between trials in the familiarization phase. Participants were told to click on the alien as quickly as possible when they saw it.

After familiarization, participants did a test phase, during which all eight word–referent pairs were tested once. The test phase started with a practice item: the word *hond* ('dog') was played and participants could choose between a picture of a dog, cat, a tree and a couch. The experimenter was allowed to provide feedback during this practice phase. There was no feedback during the actual test phase. As stated earlier, the cross-situational word learning task was part of a larger test battery. Apart from the background measures, participants also did two other statistical learning tasks that are not discussed in this chapter (see Chapter 5). The order of the tasks was counterbalanced across participants.

3.4.7 Data processing

For the off-line test phase, the practice item was removed for further analysis. For every answer it was coded whether it was correct or incorrect.

The eye-tracking data were interpolated before analysis using a Praat (Boersma & Weenink, 2019) script. When at least 1 but at most 9 consecutive samples (75 ms) in a row lacked eye-gaze information, the position of the eye in these missing samples was filled in by linear interpolation. The value of 75 ms as a maximum for a gap to be interpolated reflects a recommendation in the official Tobii manual (Tobii Pro AB, 2014). This value is chosen because it corresponds to the duration of a blink. 6.7% of the data was interpolated in this way.

After interpolation, we constructed two 1000-ms time windows. As it takes approximately 200 ms to plan an eye movement in reaction to a spoken word (Viviani, 1990), time window 1 started 200 ms after the onset of the first word. Time window 2 started 200 ms after the onset of the second word. Data points from the fixation parts of the learning trials and when the pictures were shown but the words had not yet started were removed from the data.

Two Areas of Interest (AOI's) were defined, corresponding to the two pictures that were shown on the left side and the right side of the screen during the familiarization trials. For every sample it was computed whether the participant looked at the left picture or the right picture. Trials in which more than 50% of the samples were missing (no eye-gaze data) or irrelevant (looks at the screen but not at one of the two pictures) were removed from the data (433 trials). Then, we removed all remaining missing data (31,835 samples), leaving 210,925 samples for analysis. Unfortunately, the DLD group had more missing data than the TD group (84,423 DLD; 27,891 TD), which caused an imbalance in the remaining data: 139,612 samples for the TD group and 71,313 samples for the DLD group. On average, each participant contributed data from 19.6 trials.

3.4.8 Analysis

We used the packages *lme4* (Bates et al., 2015) and *eyetracking*R (Dink & Fergusson, 2015) from the free software R (R Core Team, 2020) for data analysis. For each sample of the eye-tracking data, it was computed whether the participant looked at the target picture or the distractor picture, which depended on the word that was played at that moment. Using the eyetrackingR package, the samples were binned into 50-ms time bins. For each time bin, the proportion of looks towards the target picture was computed by dividing the number of looks towards the target by the total number of looks towards the pictures. The dependent variable was then transformed using an adjusted logit transformation.[8] In this transformed variable, a value of 0 means that a participant is looking equally often at both pictures while a positive value means s/he looks more towards the target picture. In our statistical analysis, we computed whether the proportion of looks towards the target picture depended on Group (TD/DLD), Condition (high-CD/low-CD) and Trial (1-28), keeping into account the factors Time within a trial, Age and Congruency

[8] The eyetrackingR package uses the formula Logit = log(Prop / (1 – Prop)), where Prop represents the proportion of looks towards the target image. When the value of Prop equals 0 or 1, a small value (0.1) is added: Logit = log((Prop + 0.1)/(1 – Prop + 0.1)).

(congruent vs. incongruent trials). To this end, we set up two linear mixed-effects models.

3.5 Results

3.5.1 Cross-situational word learning: accuracy

All analyses were done in R. For the off-line data, the answers to all eight test items in the test phase were taken into account. There was no missing data. Using the package *lme4*, we constructed a generalized linear mixed-effects model. Accuracy was the dependent variable, Group and Condition were between-participant predictors. Age was included as a between-participant control variable. Group and Condition were binary predictors and were coded with orthogonal contrasts: Group was coded as -½ for DLD and + ½ for TD, Condition was coded as - ½ for low-CD and + ½ for high-CD. The predictor Age was centred and scaled. The maximal model that included the main predictors and the interaction between them, random intercepts for Subject and Item as well as by-Item random slopes for Group, Condition and Age and all interactions between them resulted in a singular fit. Therefore we took out the by-item random slopes[9], resulting in the following model: Accuracy ~ Group * Condition * Age + (1 | Subject) + (1 | Item).

For answering research question 1A ('do children with DLD have more difficulty than TD children with learning word–referent pairs in an implicit cross-situational word learning task?'), the relevant effect is the main effect of Group: we expected that children with DLD learn fewer word–referent pairs on this test than TD children. For research questions 3A ('does higher contextual diversity enhance cross-situational word learning?') and 3B ('does contextual diversity impact cross-situational word learning differently in TD children than in children with DLD?'), the relevant effects are the main effect of Condition and the interaction between Condition and Group respectively: we expected more accurate responses for children in the high-CD condition compared to the low-CD

[9] This means that for the three predictors, we will not be able to generalize from our 8 specific items to a hypothetical infinite population of possible items.

condition. A significant interaction between Group and Condition would indicate that the Condition effect differed between the groups (we had no expectation about the existence or direction of such an interaction).

Our model estimates that TD children are 3.71 (95% CI: 1.73 .. 7.98) times more likely to answer an item on the test phase correctly than children with DLD: $z = 3.63$, $p = 0.0008$. Moreover, our model estimates that children in the high-CD condition score 1.67 (95% CI: 0.79 .. 3.54) times higher in the test phase than children in the low-CD condition, but this difference is not significant: $z = 1.346$, $p = 0.18$. Although the positive effect of contextual variability in the high-CD Condition was 1.1 (95% CI: 0.25 … 4.94) times stronger in the children with DLD than the TD children, this interaction between Group and Condition was not significant: $z = -0.136$, $p = 0.89$. To determine whether both groups scored higher than could be expected from chance, we also compared the performance of both groups to chance level (which was 0.25 as there were 4 possible answers on every test item). For the TD children, the estimation of the intercept, converted from log-odds to probabilities, was 0.83 (95% CI: 0.65 .. 0.94). For the DLD children, the estimate of the intercept (converted from log-odds to probabilities) was 0.49 (95% CI: 0.39 .. 0.60). For both groups, as the confidence intervals of the intercept do not include 0.25, this estimation was significantly higher than chance level. Thus, although both groups of children learned word–referent pairs in the experiment, as indicated by the above-chance performances, the children with DLD were significantly outperformed by the TD children, indicating that children with DLD have more difficulty with cross-situational word learning than TD children. We did not find evidence for or against an effect of contextual diversity on learning word–referent pairs. See Table 3.2 for the mean accuracy data, and Figure 3.5 and Figure 3.6 for a visual representation of the data.

Table 3.2 – Descriptive data accuracy off-line test phase.

	TD	DLD
low-CD	0.71 (*SD* = 0.46)	0.44 (*SD* = 0.5)
high-CD	0.74 (*SD* = 0.44)	0.55 (*SD* = 0.5)
Overall	0.73 (*SD* = 0.45)	0.49 (*SD* = 0.5)

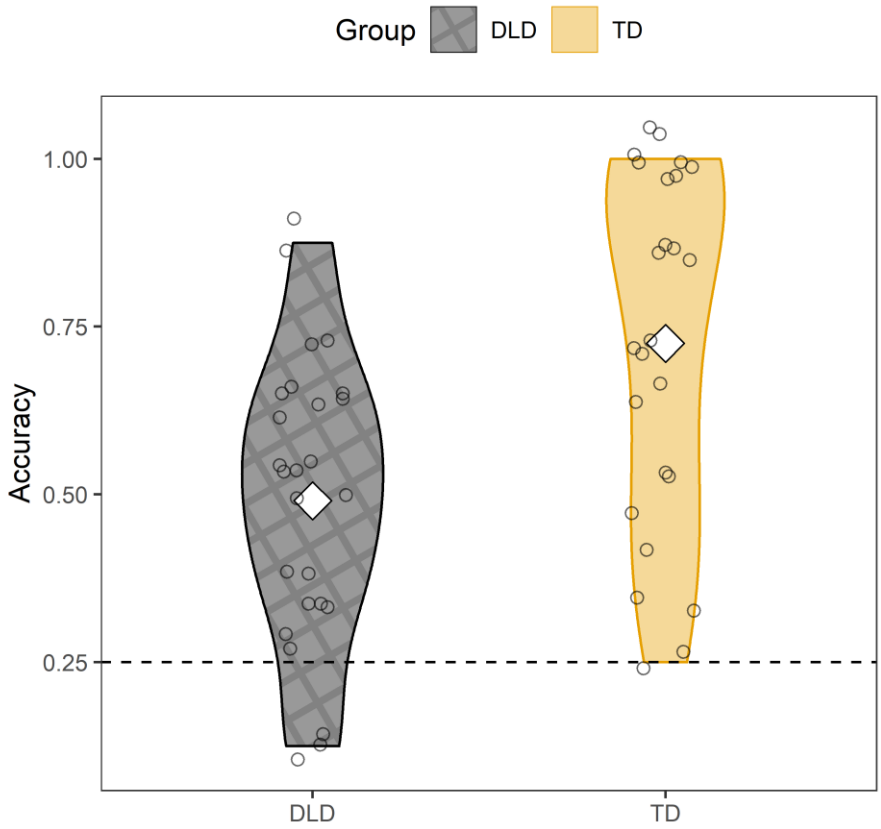


Figure 3.5 – Accuracy on the off-line test phase per Group.

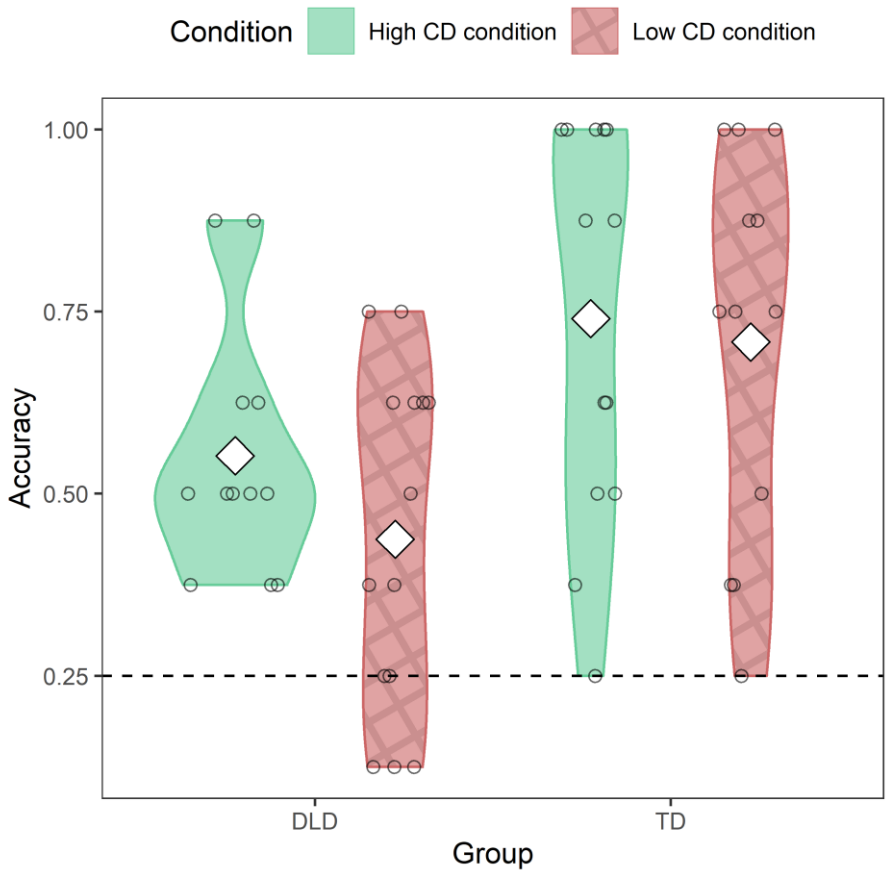


Figure 3.6 – Accuracy on the off-line test phase per Group and Condition.

3.5.2 Cross-situational word learning: eye-tracking data

We conducted two separate linear mixed-effects models for the two time windows (Word1 and Word2) to test whether the proportion of looks towards the target picture was different for children with and without DLD and whether there was an influence of Condition and Trial, taking into account the variables Time, Age and Congruency. The dependent variable was the adjusted logit transformation of the proportion of looks towards the target picture of every 50-ms time bin. Between-participant variables were Group and Condition, within-participant variables were Trial, Time, Age and Congruency. Before conducting a linear mixed-effects model, all binary variables were coded with sum-to-zero orthogonal contrasts (Group, Condition, Congruency) and all numeric variables were centred and scaled (Trial, Time, Age).

The model included all predictors and the interactions between them (except for Age), as well as random intercepts for Participant and Item. Also included were by-item random slopes (and the interactions between them) for Group and Condition as well as by-subject random slopes (and the interactions between them) for Time, Trial and Congruency. This resulted in the following model: Logit ~ Group * Condition * Time * Trial * Congruency + Age + (Group * Condition | Item) + (Time * Trial * Congruency | Participant). We are interested in the main effect of Group, Condition and Trial, as well as the interactions between Group and Condition and Group and Trial. A significant effect of Trial in the expected direction would show that children look more towards the target image as the experiment unfolds, what we interpret as an on-line learning effect. An interaction between Group and Trial would show that this on-line learning effect is different for the two groups. We are also interested whether the intercepts of the models are significant, which would indicate that children in general look more towards the target image than the distractor image.

Sanity checks and confirmatory results. In this section, we first discuss some sanity checks, and then the confirmatory results. We assume a significance criterion of $p = 0.01$. As a first sanity check, we computed whether the intercept is significantly higher than zero, which would mean that participants look more towards the target picture compared to the distractor picture. As a second sanity check, we computed whether Trial significantly influenced the proportion of looks towards the target picture, which would mean that participants look more towards the target picture as the experiment progressed (on-line learning effect). To answer our research questions 1B ('do children with DLD show weaker on-line learning than TD children during implicit cross-situational word learning?'), 3A ('does higher contextual diversity enhance cross-situational word learning?') and 3B ('does contextual diversity impact cross-situational word learning differently in TD children than in children with DLD?') we look at the effects of Group, the interaction between Group and Trial, the effect of Condition and the interaction between Group and Condition respectively. ***Word1.*** See Table 3.3 for the outcomes of the model for Word1. None of the relevant effects are significant, meaning there is no evidence that children look more towards the target picture than the distractor in general (Intercept) and whether the proportion of looks towards the target increases as the familiarization phase progresses (effect of Trial). As we do not find significant results for the sanity checks, we do not have clear evidence that the eye gaze patterns of our participants reflect on-line learning of word–referent pairs. We cannot answer our research questions, as the effects of Group, Condition, nor the interactions between Group and Trial and Group and Condition are significant. See Figure 3.7 and Figure 3.8 for a plot depicting the model predictions and actual data of the proportion of looks towards the target picture in the first time window (Word1) for the children with and without DLD and their individual differences. See Figure 3.9 for the effect of Trial across groups, and Figure 3.10 for the effect of Condition across groups.

Table 3.3 – Outcomes of the linear mixed-effects model for Word1: sanity checks and confirmatory results.

Effect	Estimate [95% CI]	t	p
Intercept	0.13 [-0.06 .. 0.32]	1.35	0.18
Trial	0.01 [-0.12 .. 0.14]	0.11	0.91
Group	0.01 [-0.35 .. 0.36]	0.04	0.97
Condition	0.15 [-0.21 .. 0.51]	0.85	0.397
*Group * Condition*	-0.15 [-0.85 .. 0.55]	-0.43	0.67
*Group * Trial*	-0.10 [-0.36 .. 0.16]	-0.76	0.45

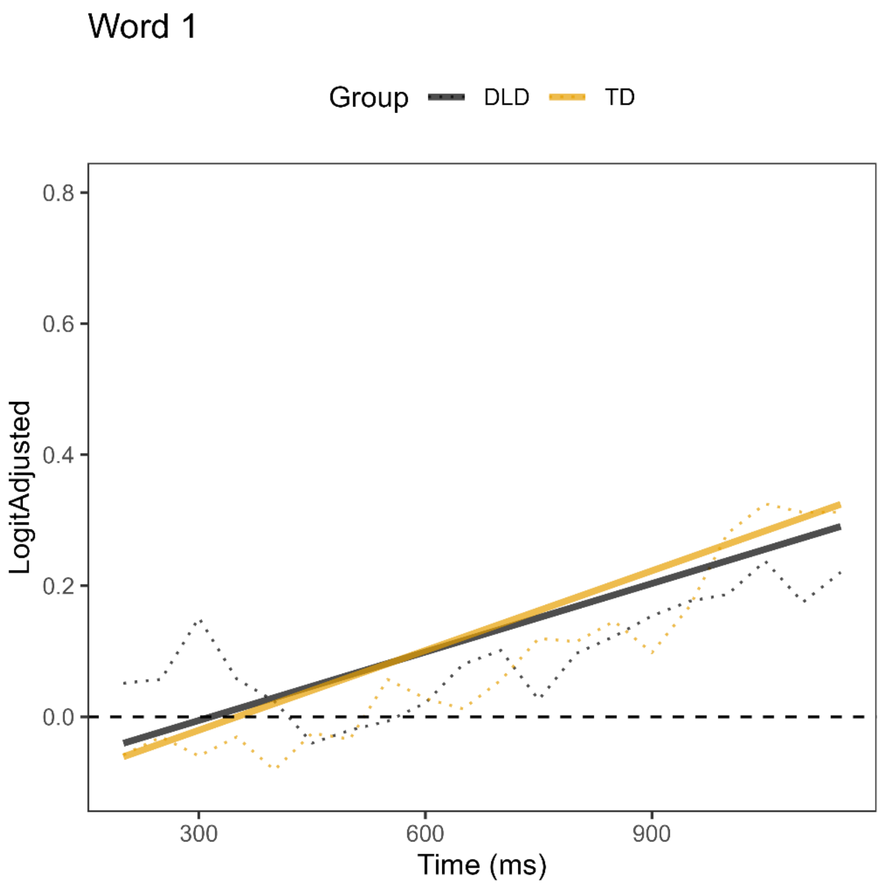


Figure 3.7 – Model prediction and actual data of looks towards the target per Group.

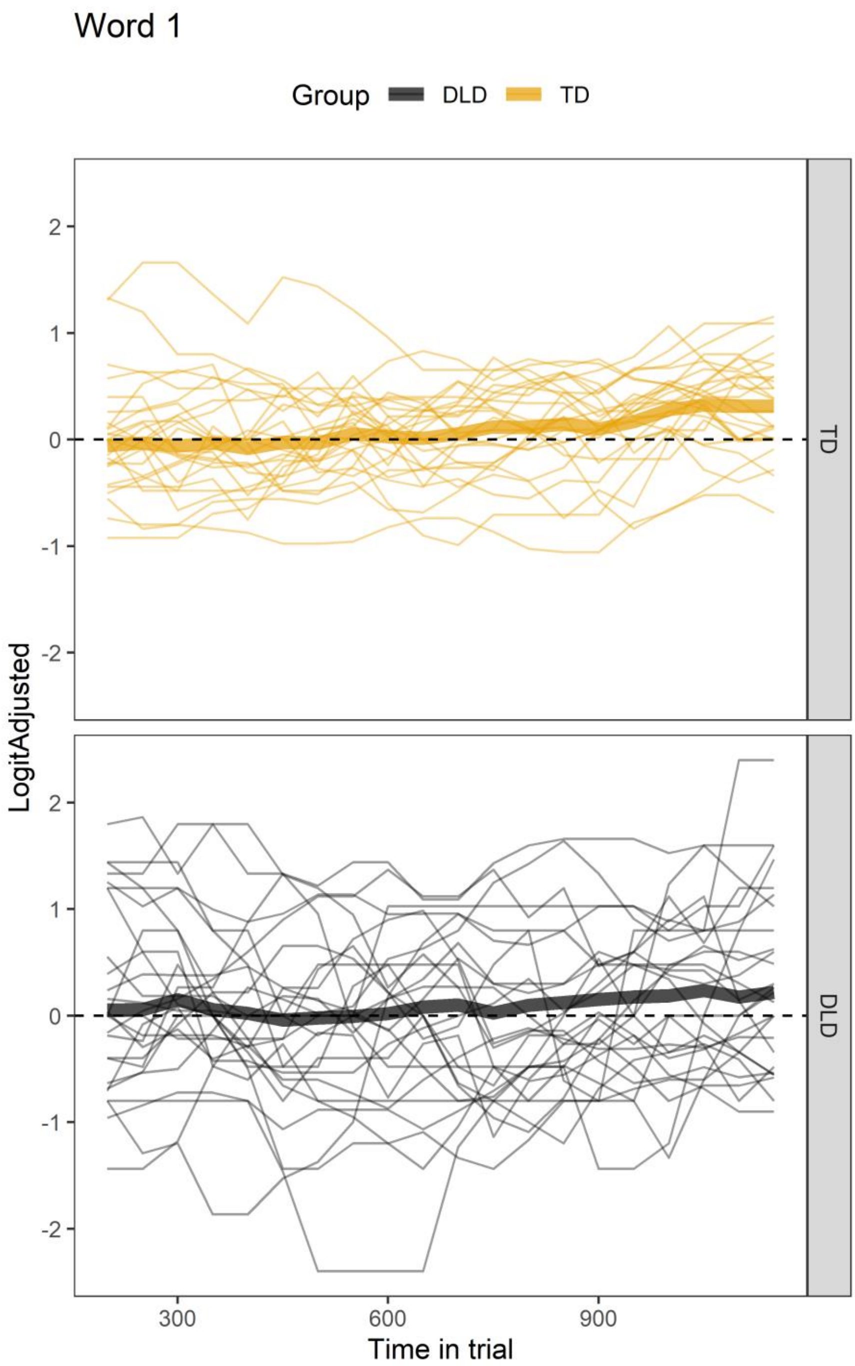


Figure 3.8 – Individual differences per Group for Word1.

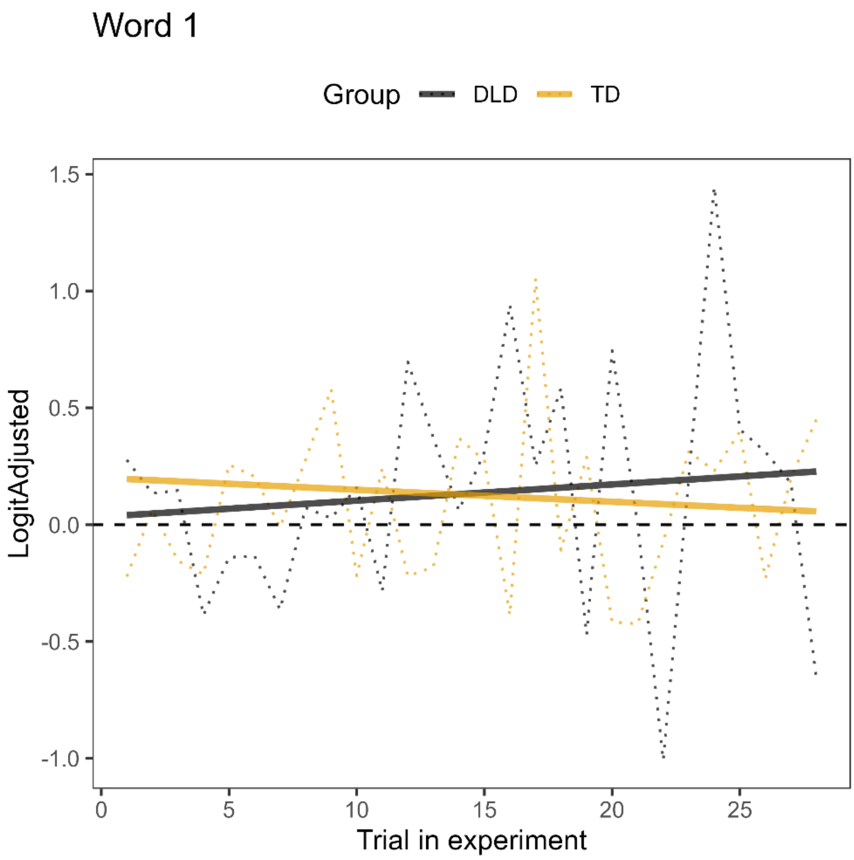


Figure 3.9 – Model prediction and actual data of the effect of Group and Trial for Word1.

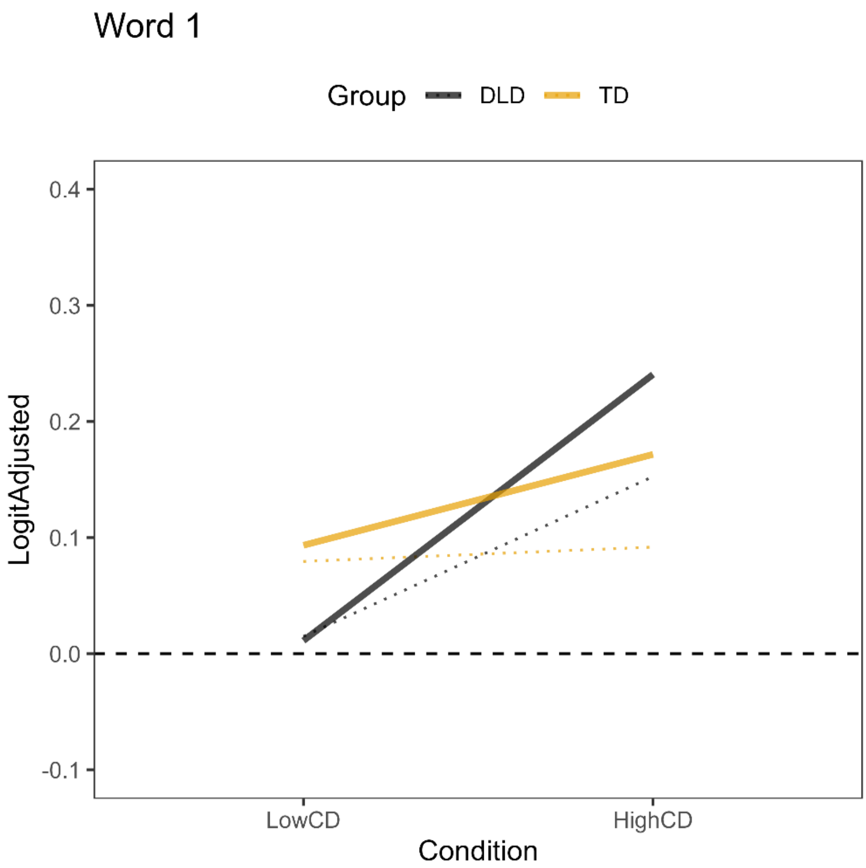


Figure 3.10 – Model prediction and actual data of the Condition effect by Group for Word1.

Word2. See Table 3.4 for the outcomes of the model for Word2. The estimate for the intercept is 0.55 logit, which is significantly different from zero (t = 4.73, p = 0.00001), indicating that children on average look more towards the target picture than the distractor picture. The other relevant effects are not significant, although we might mention that the main effect of Group and the interaction between Group and Trial approach significance. Thus, we cannot answer our research questions about the effect of group or contextual variability on implicit cross-situational learning ability. See Figure 3.11 and Figure 3.12 for a plot depicting the model predictions and actual data of the proportion of looks towards the target picture for the children with and without DLD and their individual differences. See Figure 3.13 for the effect of Trial across groups, and Figure 3.14 for the effect of Condition across Groups.

Table 3.4 – Outcomes of the linear mixed-effects model for Word2: sanity checks and confirmatory results.

Effect	Estimate [95% CI]	*t*	*p*
Intercept	0.55 [0.32 .. 0.78]	4.73	0.00001
Trial	-0.01 [-0.17 .. 0.14]	-0.18	0.86
Group	0.39 [-0.06 .. 0.83]	1.74	0.09
Condition	0.11 [-0.37 .. 0.59]	0.47	0.64
*Group * Condition*	0.01 [-0.90 .. 0.92]	0.02	0.98
*Group * Trial*	0.25 [-0.06 .. 0.55]	1.64	0.11

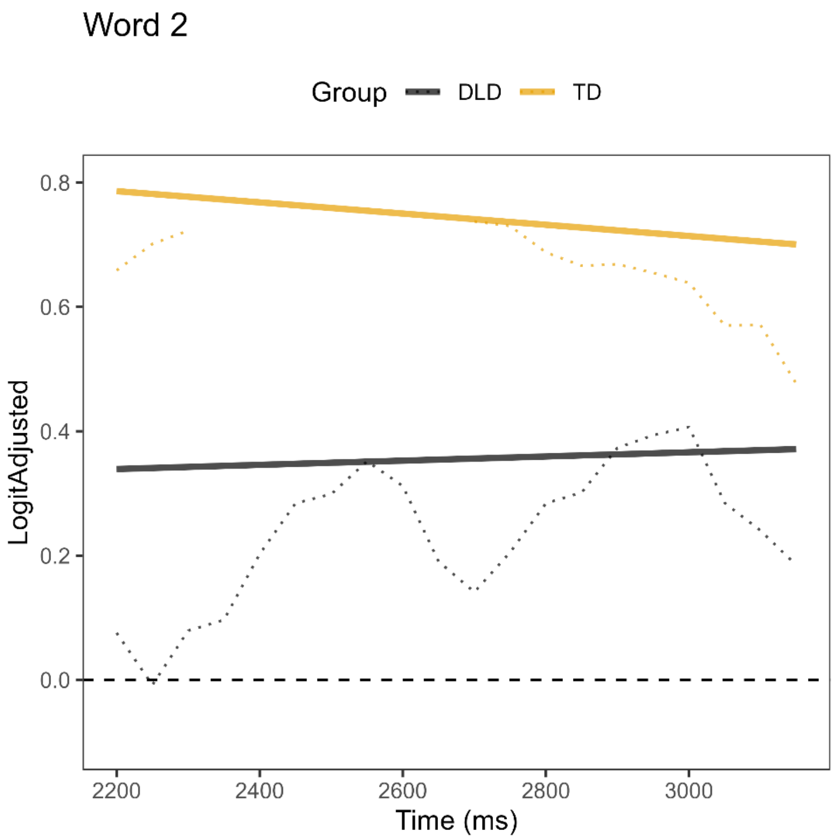


Figure 3.11 – Model prediction and actual data of looks towards the target image per Group for Word2.

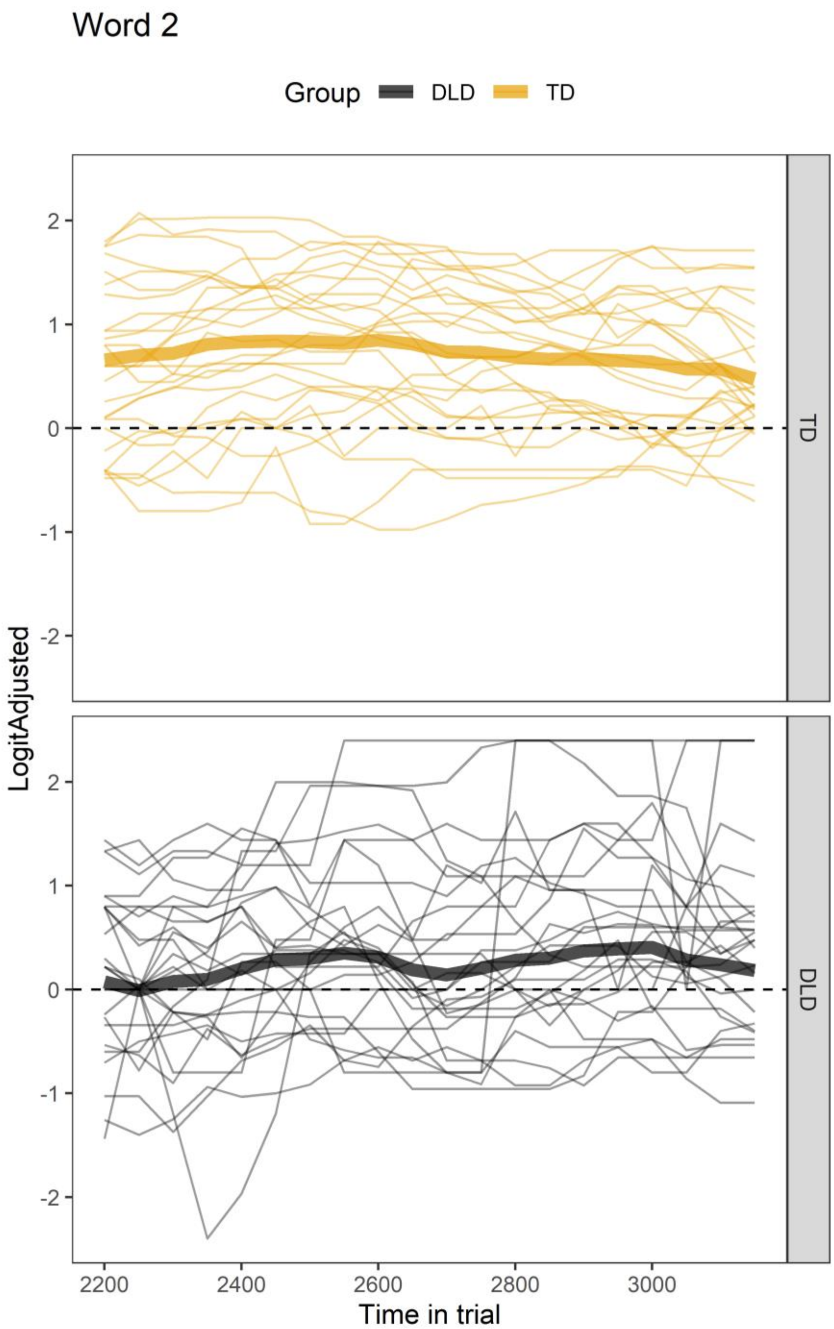


Figure 3.12 – Individual differences for Word2 per Group.

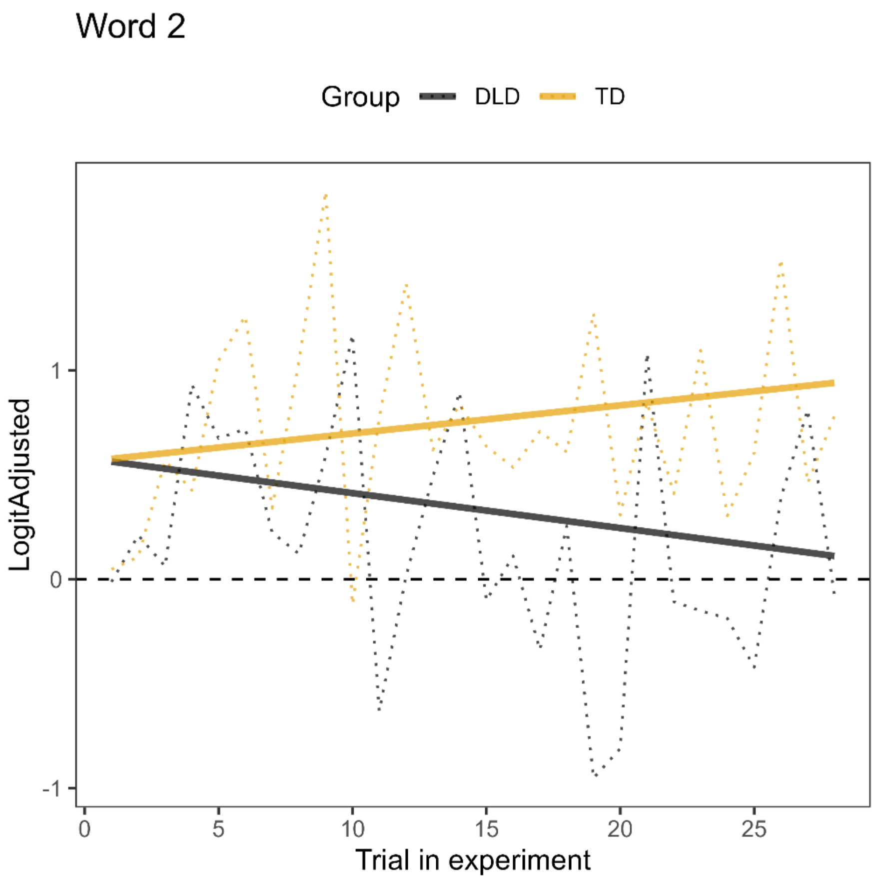


Figure 3.13 – Model prediction and actual data of looks towards the target image per Trial in the experiment for Word2.

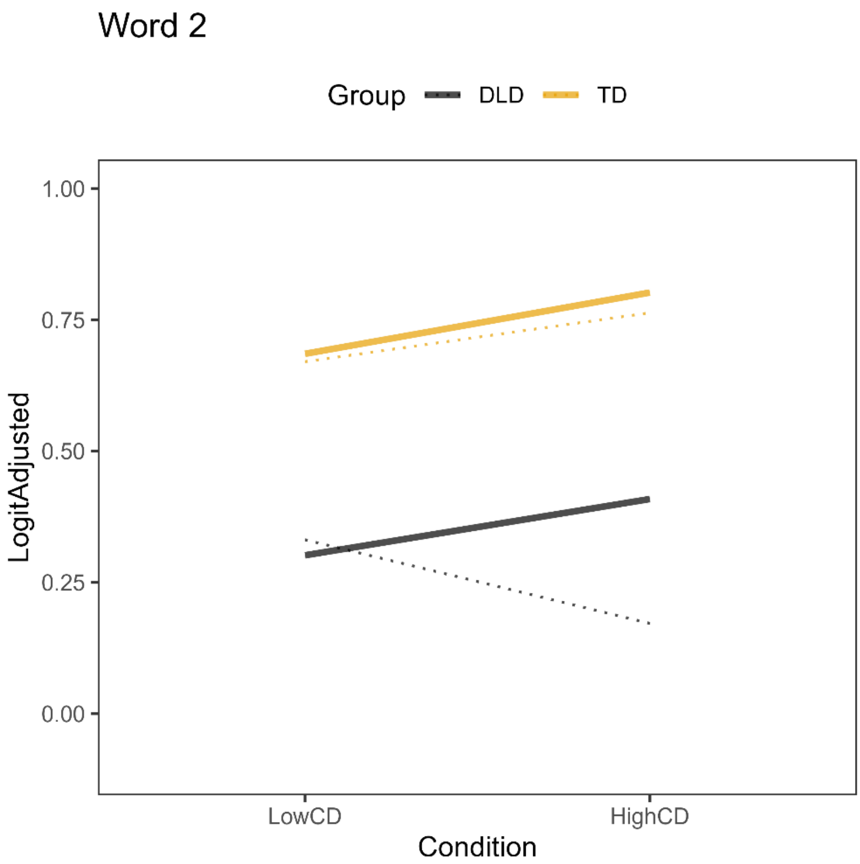


Figure 3.14 – Model prediction and actual data of the Condition effect by Group for Word2.

Exploratory results. As an exploratory analysis, we looked at the effects of Time (within trials) and Congruency (congruent/incongruent with reading order) on the proportion of looks towards the target. ***Word1.*** The children looked more towards the correct image later in a learning trial (i.e. as they have heard more of the word): estimate: 0.11, t = 2.30, p = 0.03. Moreover, children looked more towards the target picture when trials were congruent (first word refers to left image and second word to right image) than when they were incongruent (first word refers to right image and second word to left image): estimate: 0.40, t = 2.49, p = 0.015. ***Word2.*** For Word2, the effect of Time was small or non-existent (95% CI: -0.10 .. -0.08, t = -0.18, p = 0.86). Again, children's looking times were influenced by Congruency: estimate: 0.49, t = 2.22, p = 0.03. It could be the case that time within a trial and the order of the words and pictures influences looking behaviour.

3.5.3 Regression analyses

To answer research question 2 ('is cross-situational word learning ability related to lexical-semantic skills in children with DLD?'), we performed regression analyses to investigate the relationship between cross-situational word learning ability on the one hand and existing lexical-semantic knowledge on the other hand. Dependent variables were the different measures of lexical-semantic knowledge, predictor variables were off-line and on-line measures of cross-situational word learning and several control measures of cognitive abilities, as well as age and SES. All variables were centred and scaled before analysis.

We constructed a principal component analysis (PCA) in R on the measures of non-verbal intelligence, digit span forwards, digit span backwards and non-word repetition to reduce the number of predictor variables. This resulted in four component scores, of which we decided to use the first three as they together explained 95% of the variance. See Table 3.5 for the standardized loadings of the component scores after varimax rotation. The scores represent phonological processing (digit span forwards, non-word repetition), non-verbal intelligence (Raven) and verbal working memory (digit span backwards) and were saved and used for as predictor scores in further analyses. We computed the correlations between the predictor variables (See Table 3.6).

Table 3.5 – Standardized loadings of the component scores from the PCA.

	Component 1 (phonological processing) 46%	Component 2 (non-verbal intelligence) 27%	Component 3 (verbal working memory) 26%
Digit span forwards	0.93	-0.22	0.03
Digits span backwards	0.04	0.21	0.98
Non-word repetition	0.95	0.13	0.02
Non-verbal intelligence	-0.05	0.97	0.22

Table 3.6 – Correlations between predictor variables. Marked in grey are significant positive correlations.

	CSWL (on-line)	Comp. 1	Comp. 2	Comp. 3	Age	SES
CSWL (off-line)	r = 0.19 p = 0.35	r = -0.28 p = 0.17	r = 0.38 p = 0.0	r = -0.01 p = 0.97	r = -0.06 p = 0.77	r = -0.15 p = 0.45
CSWL (on-line)		r = -0.26 p = 0.20	r = 0.41 p = 0.04	r = -0.05 p = 0.797	r = -0.3 p = 0.13	r = 0.11 p = 0.60
Comp. 1			r = 0 p = 1	r = 0 p = 1	r = 0.11 p = 0.58	r = 0.05 p = 0.83
Comp. 2				r = 0 p = 1	r = -0.2 p = 0.33	r = 0.01 p = 0.96
Comp. 3					r = 0.12 p = 0.57	r = -0.3 p = 0.13
Age						r = 0.15 p = 0.46

Notes: Component 1 = phonological processing, component 2 = non-verbal intelligence, component 3 = verbal working memory

We constructed four separate linear regression models for the four dependent variables, which are discussed one by one below. CSWL off-line represent the average accuracy on the test phase of the CSWL task, while CSWL on-line represents the mean proportion of looks towards the target picture during the familiarization phase of the CSWL task.

Passive vocabulary. The linear model with passive vocabulary as the dependent variable as a whole was not significant (F = 0.6801, p = 0.69, adjusted R^2 = -0.098), meaning that the full model did not predict the dependent variable better than a null model without any predictors. See Table 3.7 for the contributions of the individual predictors. None of the predictors contributed significantly to variance in passive vocabulary size in the children with DLD.

Table 3.7 – Results from the linear model of passive vocabulary.

Predictor	Estimate (log odds) [95% CI]	Std. error (log odds)	t	p
Age	0.15 [-0.33 .. 0.63]	0.23	0.65	0.52
SES	0.16 [-0.33 .. 0.64]	0.23	0.68	0.51
Component 1 (phonological processing)	0.25 [-0.23 .. 0.74]	0.23	1.1	0.28
Component 2 (non-verbal intelligence)	0.14 [-0.39 .. 0.67]	0.25	0.56	0.58
Component 3 (verbal working memory)	0.11 [-0.37 .. 0.58]	0.22	0.644	0.64
CSWL off-line	0.22 [-0.28 .. 0.73]	0.24	0.366	0.37
CSWL on-line	-0.19 [-0.71 .. 0.34]	0.25	-0.75	0.46

Active vocabulary. The linear model with active vocabulary as the dependent variable as a whole was not significant (F = 1.196, p = 0.35, adjusted R^2 = 0.05). See Table 3.8 for the contributions of the individual predictors. None of the predictors significantly contributed to the children's active vocabulary score.

Table 3.8 – Results from the linear model of active vocabulary.

Predictor	Estimate (log odds) [95% CI]	Std. error (log odds)	*t*	*p*
Age	0.22 [-0.22 .. 0.67]	0.21	1.05	0.31
SES	0.37 [-0.09 .. 0.82]	0.22	1.70	0.11
Component 1 (phonological processing)	0.27 [-0.18 .. 0.71]	0.21	1.25	0.23
Component 2 (non-verbal intelligence)	0.13 [-0.36 .. 0.62]	0.23	0.57	0.57
Component 3 (verbal working memory)	0.04 [-0.40 .. 0.47]	0.21	0.18	0.86
CSWL off-line	0.09 [-0.39 .. 0.56]	0.22	0.38	0.71
CSWL on-line	-0.12 [-0.61 .. 0.37]	0.23	-0.52	0.61

Word categories. The linear model with word categories score as the dependent variable as a whole was not significant (F = 1.827, p = 0.14, adjusted R^2 = 0.19). See Table 3.9 for the contributions of the individual predictors. None of the predictors significantly contributed to the children's word categories score.

Table 3.9 – Results from the linear model of word categories score.

Predictor	Estimate (log odds) [95% CI]	Std. error (log odds)	*t*	*p*
Age	-0.37 [-0.78 .. 0.05]	0.20	-1.87	0.08
SES	0.04 [-0.38 .. 0.46]	0.20	0.19	0.85
Component 1 (phonological processing)	-0.18 [-0.59 .. 0.24]	0.20	-0.90	0.38
Component 2 (non-verbal intelligence)	0.32 [-0.13 .. 0.78]	0.22	1.49	0.15
Component 3 (verbal working memory)	0.18 [-0.22 .. 0.59]	0.19	0.96	0.35
CSWL off-line	0.22 [-0.22 .. 0.65]	0.21	1.04	0.31
CSWL on-line	-0.10 [-0.55 .. 0.35]	0.21	-0.48	0.64

Word associations. The linear model with word associations score as the dependent variable as a whole was not significant (F = 0.799, p = 0.59, adjusted R^2 = -0.06). See Table 3.10 for the contributions of the individual predictors. None of the predictors significantly contributed to variance in word associations score in the children with DLD.

Table 3.10 – Results from the linear model of the word associations score.

Predictor	Estimate (log odds) [95% CI]	Std. error (log odds)	*t*	*p*
Age	-0.05 [-0.53 .. 0.42]	0.23	-0.24	0.82
SES	0.24 [-0.24 .. 0.72]	0.23	1.05	0.31
Component 1 (phonological processing)	0 [-0.47 .. 0.48]	0.23	0.02	0.98
Component 2 (non-verbal intelligence)	0.22 [-0.30 .. 0.74]	0.25	0.88	0.39
Component 3 (verbal working memory)	-0.23 [-0.69 .. 0.23]	0.22	-1.03	0.32
CSWL off-line	0.05 [-0.45 .. 0.55]	0.24	0.22	0.83
CSWL on-line	-0.38 [-0.90 .. 0.13]	0.24	-1.57	0.13

3.6 Discussion

The current study aimed to investigate implicit cross-situational word learning in children with and without DLD and its relation to lexical-semantic knowledge. We will discuss the results per research question in the sections below.

3.6.1 RQ1: Are children with DLD less proficient in cross-situational word learning?

Results from the analysis of the off-line test phase show that both our groups were able to pick up the mappings between novel objects and novel pictures, while they had not received instructions to do so. This indicates that children with and without DLD can use statistical learning mechanisms to link words and referents implicitly. However, as our children with DLD performed significantly lower than our TD children (p = 0.0008), we can conclude that children with DLD likely are not able to profit from statistical learning to the same extent as children without DLD do. These results are in line with the findings of McGregor et al. (2022)

and Ahufinger et al. (2021). The latter also report above-chance performance for both TD children and children with DLD, but poorer performance in the last group in a more explicit learning condition. Our study extends this finding to implicit cross-situational word learning.

We also aimed to measure the process of learning using eye-tracking. We expected that children would start to look more towards the target image as the experiment progressed, reflecting learning of word–referent pairs during the exposure phase of the experiment. Moreover, we expected to find group differences in looking behaviour. One finding seems to reflect on-line learning: the intercept for the model of the second word in a learning trial was significant, showing that children have a preference for the target picture as opposed to the distractor picture, corresponding to the finding of above chance on the off-line test phase. This is an extension of the eye-tracking results of Ahufinger et al. (2021), who did not report any evidence for a preference for the target image. However, the remaining predictors did not significantly influence looking behaviour. The effect of Trial was not significant for the first word or the second word, meaning we have no evidence for an on-line learning effect across trials. Since neither the main effect of Group nor the interaction between Group and Trial was significant, we have no evidence that children with DLD look less often towards the target picture in general or that they show less strong on-line learning. Exploratory analyses might indicate that time within a trial and the congruency of the order of the words and pictures (congruent with reading order or incongruent) influenced the proportion of looks towards the target, but we cannot draw any conclusions about exploratory findings.

As can be seen in Figure 3.8 and Figure 3.12, the amount of individual variation is large, especially within the DLD group. Moreover, as discussed in § 3.4.7, the contribution of data points between the two groups is highly skewed: the TD children provided many more data points than the children with DLD. In Appendix 1 and Appendix 2, graphs are provided that show the number of data points per group, split up for the predictors Time, Condition, Congruency and Trial. Besides the overall imbalance between the groups, data for the predictors Condition, Congruency and Trial are also skewed for the DLD group. While for the TD children the data is roughly equally divided, for the children with DLD there is more data for the low-CD condition, for the incongruent trials and the earlier trials in the experiment than there is for their counterparts. These imbalances are caused by the large number of missing data in our DLD group, but also relatively many 'irrelevant' looks (looks at the screen but outside the AOIs; 39,677 samples in the DLD data versus 12,461 samples in the TD data). It could be the case that the eye-tracker worked less efficiently for these children, but is likely that they looked less well at the screen overall. This could be related to attention difficulties, which have been established in children with DLD (for a review, see Smolak et al., 2020), and McGregor et al. (2022) report that sustained attention predicts cross-situational word learning ability in children with DLD. The skewness of the data and the large individual variation possibly weakened statistical power, which could partly explain the absence of significant effects. Future studies should aim to test larger groups of participants.

One could argue that learning could have faltered at the level of phonology for the children with DLD. As children with DLD are shown to have difficulty with phonological short-term memory and seem to store less specified phonological representations (Mainela-Arnold et al., 2010), it might be hard for them to disentangle the new words in their memory, resulting in poorer learning. To reduce the chance that children would confuse the words, we chose to have more variation in phonological structure then is often implemented: the words in our experiment have different (simple) phonological structures (CVC, CVCV, CVCVC) and every word starts with a different consonant. Still, as Bogaerts et al. (2021) argue, it would be fruitful for future studies to set up experiments that can show a contrast between impaired statistical learning and intact performance on a task that does not entail statistical learning.

3.6.2 RQ2: Is cross-situational word learning ability related to lexical-semantic skills in children with DLD?

We expected to find that cross-situational word learning ability significantly contributes to lexical-semantic knowledge in children with

DLD. Besides segmenting words from running speech, tracking the co-occurences between auditory words and visual referents contributes to gaining lexical-semantic knowledge. For children with DLD, it might be the case that this type of implicit word learning works less eficiently and hampers lexical-semantic development. However, as none of the multiple linear regression models we conducted was significant, we cannot conclude anything about the relation between implicit cross-situational word learning and existing lexical-semantic knowledge in children with DLD, nor the influence of age, SES, phonological processing, non-verbal intelligence and verbal working memory. Future studies, besides testing larger participant groups, could investigate this relationship by setting up longitudinal experiments.

Previous work has shown a relationship between cross-situational word learning ability and vocabulary size in young TD children (22-66 months; Vlach & DeBrock, 2017). Kemény and Lukács (2021) report a significant independent contribution of probabilistic statistical learning ability (weather prediction task) to vocabulary size in TD children, while short-term memory did not independently contribute to vocabulary. However, in their children with DLD, this pattern was reversed: short-term memory independently contributed to vocabulary size, but statistical learning ability did not. The authors interpret the results as indicating that different cognitive abilities underlie lexical development in TD children and children with DLD, although it is important to note that this interpretation is based on a p-value comparison. McGregor et al. (2022) report that vocabulary is a predictor of cross-situational word learning ability, and that this relationship is stronger in TD children compared to children with DLD, based on a relative importance analysis. It could be the case that children with DLD compensate for less efficient statistical learning mechanisms by depending more on, for example, declarative learning, which might explain why we did not find a significant relationship between cross-situational word learning and lexical-semantic knowledge in our group of children with DLD. Unfortunately, we were not able to compare the contribution of cross-situational word learning to vocabulary between children with and without DLD, as the TD children in our experiment were not tested on lexical-semantic skills.

3.6.3 RQ3: Does high contextual diversity enhance cross-situational word learning?

We manipulated contextual diversity between subjects to investigate whether variability in the learning environment would affect cross-situational word learning in children with and without DLD. Although performance was higher in the condition with higher contextual diversity on average, there was no significant effect of condition, nor a significant interaction between condition and group, and thus we cannot answer the question whether variability in the learning environment influences cross-situational word learning in children with and without DLD. The eye-tracking data also did not reveal evidence for a difference in on-line learning for the two conditions.

3.7 Concluding remarks and future research

Our study shows that children with DLD are less proficient when learning word meanings based on cross-situational statistics in an implicit task. If utilizing contexts with different amounts of referential uncertainty by implicitly tracking co-occurrences between words and visual referents works less efficiently in children with DLD, this could hamper the acquisition of vocabulary. Although the relationship between cross-situational word learning and existing lexical knowledge requires more investigation, our study contributes to our knowledge of different types of statistical learning in children with DLD. The cross-situational word learning paradigm aims to mimic real-life situations with referential uncertainty. However, it is far from realistic. Zhang et al. (2021) investigated naturalistic cross-situational word learning in children who are playing with toys. Future research could compare this naturalistic cross-situational word learning between children with and without DLD.

Chapter 4

School-aged children learn novel categories on the basis of distributional information

This chapter is a slightly modified version of the published article:

Broedelet, I., Boersma, P., & Rispens, J. (2022). School-aged children learn novel categories on the basis of distributional information. *Frontiers in Psychology*, 12. https://doi.org/10.3389/fpsyg.2021.799241

R scripts, data and materials are available on FigShare:
https://doi.org/10.21942/uva.c.5015285.v1

Abstract

Categorization of sensory stimuli is a vital process in understanding the world. In this chapter we show that distributional learning plays a role in learning novel object categories in school-aged children. An 11-step continuum was constructed based on two novel animate objects by morphing one object into the other in 11 equal steps. Forty-nine children (7-9 years old) were subjected to one of two familiarization conditions during which they saw tokens from the continuum. The conditions differed in the position of the distributional peaks along the continuum. After familiarization it was tested how the children categorized the stimuli. Results show that, in line with our expectations, familiarization condition influenced categorization during the test phase, indicating that the frequency distribution of tokens in the input had induced novel object category formation. These results suggest that distributional learning could play an important role in categorizing sensory stimuli throughout life.

4.1 Introduction

The world around us is incredibly complex. We need to form mental categories in order to make sense of the sensory information we perceive, which allow us to recognize and distinguish different objects, for example distinguishing a knife from a screwdriver. *Categorical perception* reflects the phenomenon that aforementioned mental categories influence how we process information: differences between objects from the same category are less important and thus more difficult to process than differences between objects from distinct categories (Collins & Olson, 2014; Harnad, 1990). Experimental studies with participants across the lifespan have demonstrated categorical perception of phenomena such as familiar objects (e.g. Newell & Bülthoff, 2002, in adults), colors (e.g. Witzel & Gegenfurtner, 2016, in adults), faces (Altvater-Mackensen et al., 2017, in infants) and speech sounds (e.g. Liberman et al., 1957, in adults; Maye et al., 2002, in infants; Vandermosten et al., 2019, in children). In the study of Newell and Bülthoff (2002), adult participants showed categorical perception of different familiar objects like bottles, glasses and lamps in adults. Linear continuums of three-dimensional visual stimuli were constructed, e.g. a transformation of a wine glass to a beer glass in 11 equal steps. Participants perceived these continuums as categorical rather than continuous: results from an identification task showed that there was a clear point where the object was no longer a wine glass, but a beer glass. Moreover, the experiment showed better discrimination of two tokens that surround that boundary (between-category discrimination) than of two tokens within a category (within-category discrimination).

How do humans build such mental categories? Top-down information such as linguistic labels play a role in forming object categories (Plunkett et al., 2008; Waxman & Gelman, 2009), but bottom-up learning, that is learning from low-level auditory and/or visual features without prior knowledge of the category label, is important for category formation as well. Research suggests that *statistical learning*, a learning mechanism that underlies the extraction of regularities from sensory input (Siegelman, Bogaerts, & Frost, 2017) contributes to bottom-up category learning, by detecting the similarities between different entities (Sloutsky, 2003). In statistical learning research, it has been shown that infants, children and adults extract regularities from the environment in the linguistic and the visual domain (e.g. Sherman et al., 2020). For example, infants are able to track the co-occurrence of shapes when exposed to complex scenes (Zsef Fiser & Aslin, 2002). In more recent work, Wu et al. (2010) and Wu et al. (2011) showed that infants are sensitive to co-occurring visual features and that they can use this information to learn about object integrity.

A specific type of statistical learning that is important for category formation is *distributional learning*. Distributional learning is defined as learning from exposure to the relative frequency of stimuli in the environment. Maye et al. (2002, 2008) proposed the hypothesis that distributional learning underlies the formation of phonetic categories. In their experiment, 6- to 8-month-old infants were familiarized with a speech sound continuum. Infants were subjected to one of two possible familiarization conditions. For infants in the bimodal condition, sounds from the near endpoints of the continuum were presented most frequently, whereas for infants in the unimodal condition, sounds from the middle part of the continuum were most frequent. After training, the bimodally trained infants turned out to be able to distinguish the endpoints of the continuum better from each other than the unimodally trained infants. These experiments therefore showed evidence that distributional information helps infants to acquire the sound categories that are relevant for their native language.

After the studies of Maye et al. (2002, 2008), distributional learning of phonetic categories in infants has also been found by Wanrooij et al. (2014). Moreover, evidence is reported for 8–9-year-old children (Vandermosten et al., 2019) and adults (e.g. Hayes-Harb, 2007). This accumulated evidence supports the plausibility of the findings in distributional learning studies (ter Schure et al., 2016). The study of Hayes-Harb (2007) suggests that distributional learning mechanisms also play a role when adults learn new phonetic contrasts in a second language. Vandermosten et al. (2019) found that also school-aged children can learn a new phonetic contrast based on distributional cues and that children with dyslexia seem to be less sensitive to those cues.

Infants, children and adults are thus able to build *phonetic* categories based on the distributional regularities in the input. But is this specific to phonetics or does it generalize to other cognitive domains? For example, does distributional learning also support the formation of *visual* categories? Altvater-Mackensen et al. (2017) investigated whether infants are sensitive to distributional cues when learning about new faces in an EEG study. In a design similar to Maye et al. (2002), a continuum that morphed from one female face to another was constructed and bimodal and unimodal familiarization conditions were compared. Results showed that infants in the bimodal group are better at discriminating two faces from the endpoints of the continuum compared to participants in the unimodal group, indicating that they form two categories. In another study, Junge et al. (2018) applied the research design of Maye et al. (2002) to novel object category learning. Six to 8-month-old infants were familiarized with exemplars from an 8-step continuum of two novel objects. Again, it was shown that infants that are subjected to the bimodal condition have stronger discrimination than infants that are subjected to the unimodal condition. These studies suggest that distributional learning is a domain-general learning mechanism underlying the categorization of auditory as well as visual stimuli, at least in infancy. It is yet unknown whether visual distributional learning plays a role in novel object categorization in older children as well. As the visual environment is endlessly variable and everchanging it is probable that distributional learning plays a role in learning about new object categories beyond the age of infancy.

Previous evidence thus suggests that distributional learning plays a role in the formation of categories of sounds, faces and novel objects. Distributional learning research predominantly focuses on learning in infants. Therefore it is presently unknown whether visual distributional learning plays a role in categorizing novel objects in older children. Research on the formation of phonetic categories shows that distributional learning mechanisms play a role throughout life. In the current study, we investigated whether bottom-up distributional learning contributes to categorizing novel visual stimuli in school-aged children.

Importantly, the conventional unimodal-versus-bimodal experimental design used in distributional learning studies has been criticized recently as it appears to contain a confounding factor (Wanrooij et al., 2015). Namely, when bimodal and unimodal distribution conditions are compared, not only the number of peaks differ between conditions but also the dispersion (or spreading) of the exemplars along the continuum. Specifically, in the usual distributional learning designs, the standard deviation between the stimuli in the bimodal condition is higher than that between the stimuli in the unimodal condition, which could result in better discrimination for the bimodal group. Wanrooij et al. (2015) constructed bimodal and unimodal distributions that were controlled for dispersion to test this prediction and found (when comparing the null hypothesis with four other plausible hypotheses) that it is likely that people in the bimodal condition cannot discriminate endpoint tokens better than people in the unimodal condition. The authors state that previous research on distributional learning might be unreliable because of the confounding factor of dispersion. Therefore it is important to take this factor into account.

In the current chapter we adapted the design of Chládková et al. (2022), who compared learning on two bimodal familiarization conditions (instead of comparing unimodal and bimodal conditions) in adult second language learners of Spanish. Using this design, they did not test whether participants learned two categories or one broad category, but whether they learned two different sets of two categories depending on the location of the distributional peaks in the continuum. In the two bimodal familiarization conditions the peaks were located at different points in the continuum. We hypothesized that children learn that two tokens that fall within one distributional peak belong to one category, while tokens that fall into different distributional peaks belong to two different categories. As those peaks were different in the two conditions, we were able to test whether children categorize the same stimuli differently dependent on condition.

Previous evidence suggests that infants, children and adults are sensitive to auditory and visual statistical information, and on that basis we hypothesized that school-aged children are also able to learn novel object categories based on the distributional properties of the input.

4.2 Method

4.2.1 Participants

Fifty children (23 females, 27 males) were recruited via two primary schools in the Netherlands. One child was excluded from analysis because of the diagnosis of dyslexia. Their ages varied between 7;6 (years;months) and 9;9 (*M* = 8;6, *SD* = 1;1). All children were native speakers of Dutch and had been brought up monolingually. They did not have any hearing difficulties, serious visual problems, nor a diagnosis of autism spectrum disorder, AD(H)D, learning difficulties, developmental dyslexia or any other language-based disorders. Ethical approval for the experiment was obtained from the Ethical Committee of the faculty of Humanities of the University of Amsterdam. The caretakers of the children filled in an informed consent form prior to their participation. Each child was randomly assigned to one of the two familiarization conditions. As the exclusion of one child resulted in an odd number of participants, 25 children did Condition 1, while 24 did Condition 2.

4.2.2 Stimuli and design

A continuum ranging from one visual object to another in 10 equal steps was constructed using the Sqirlz 2.1 software (Xiberpic.com). The endpoint stimuli (photos of two toys from Giant Microbes www.giantmicrobes.com) were copied from Junge et al. (2018) with their permission. We constructed 9 intermediate pictures to arrive at an 11-point continuum (as opposed to the 8-point continuum that was used by Junge et al. (2018) to adapt the design of Chládková et al. (2022).

The familiarization phase consisted of 12 blocks in which 24 stimuli were presented (288 stimuli in total). Two conditions were developed following a between-participant design. In Condition 1 (see Figure 4.1, orange curve), tokens 3 and 7 were most frequent, while in

Condition 2 (see Figure 4.1, blue curve) tokens 5 and 9 were most frequent. The frequencies of the different tokens were one, two, three or four times per block, resulting in a total occurrence of 12, 24, 36 or 48 times after 12 blocks (see also Figure 4.1 for the frequencies of the different stimuli). The peaks reflected the categories in the continuum. Three of the tokens were used to test categorization after the familiarization phase, and therefore all occurred equally frequently in both familiarization conditions: token 6 (referred to as the standard (S)), token 4 (referred to as deviant 1 (D1)) and token 8 (referred to as deviant 2 (D2)). In Condition 1, S and D2 belong to one distributional peak, while in Condition 2, S and D1 belong to one distributional peak. If the distributional properties of the input affect categorization of visual stimuli from a continuum, tokens from a distributional peak should be perceived as being more similar compared to tokens from two different peaks. The stimuli were presented in a random order, one by one against a dark grey background. Each stimulus was presented for 800 ms with an interstimulus interval of 200 ms (based on Arciuli & Simpson, 2011 and Turk-Browne et al., 2005). The familiarization phase contained two randomly placed filler stimuli per block (24 in total). The filler stimuli functioned as a cover task; they moved about the screen and participants were asked to click on them as fast as they could.

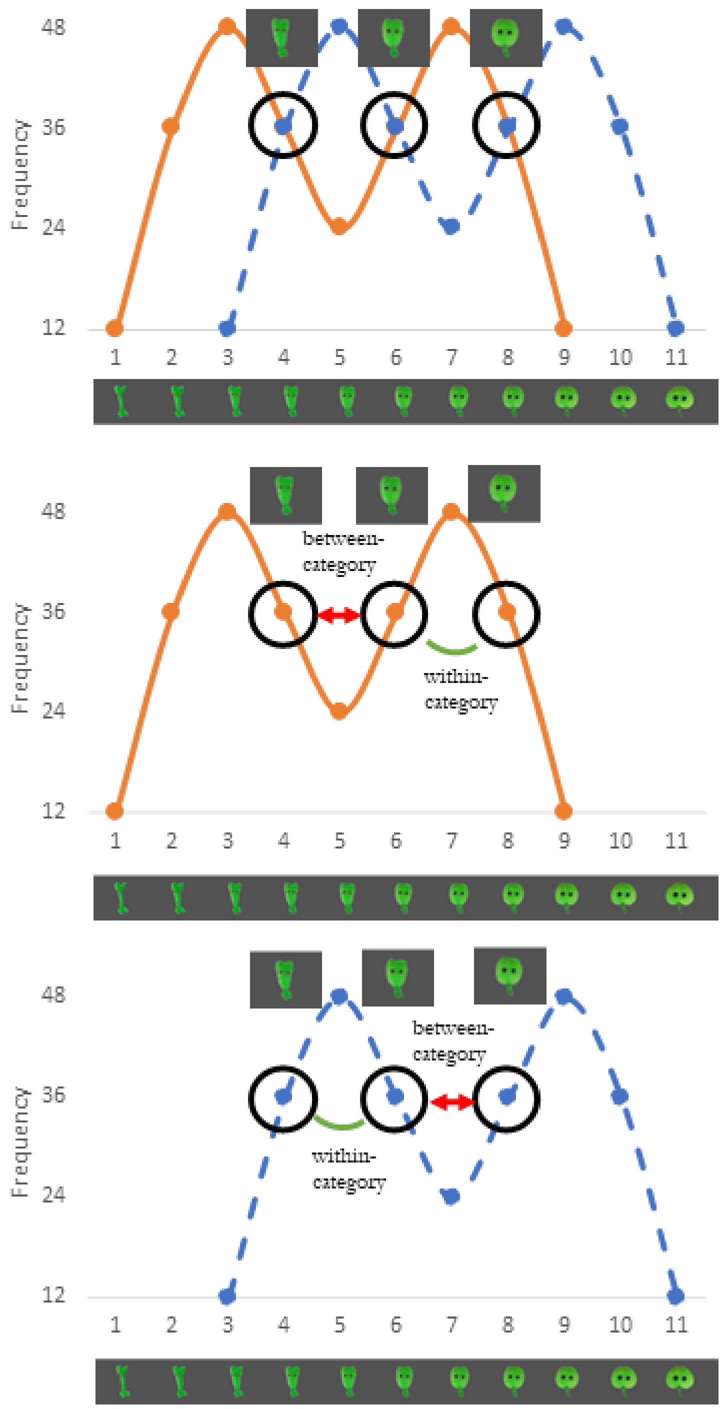

Frequency
48
36
24
12
1
2
3
4
5
6
7
8
9
10
11
between-category
within-category
between-category
within-category

Figure 4.1 - Design of the familiarization conditions of the current study. In Condition 1 (orange curve), S and D2 belong to one distributional peak, while in Condition 2 (blue curve), S and D1 belong to one distributional peak.

The test phase consisted of a practice question, eight test questions and four filler questions. We constructed AXB questions to test for categorization. In each test item, all three stimuli were presented simultaneously. S was shown in the upper part of the screen and D1 and D2 were shown below a white stripe (see Figure 4.2, left). All test questions were similar, but the position of D1 and D2 was counterbalanced across trials. Participants had to choose which of the two stimuli below the stripe looked more like the one above. The filler questions (Figure 4.2, right) were added to make the test phase less repetitive. The practice and filler questions were the same as the test questions, except that the filler stimuli from the familiarization phase were used. The test phase was the same for participants from both conditions, but the predictions were different: participants from Condition 1 were expected to have categorized S and D2 together and thus should pick stimulus D2 more often than participants who did Condition 2, who were expected to have categorized S and D1 together and thus pick D1 more often. In other words, D2 was the target answer for participants in Condition 1, while D1 was the target answer for participants in Condition 2.

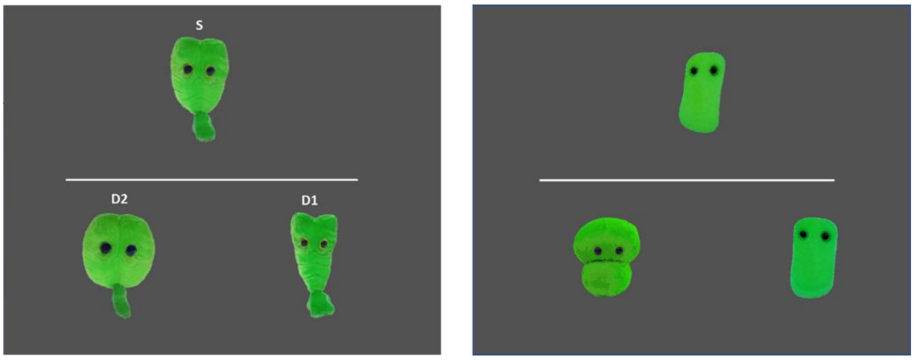


Figure 4.2 – Example of a test trial and a filler/practice test trial. Participants had to choose which of the two lower pictures was a better match for the upper picture.

4.2.3 Procedure

The experiment was run in E-Prime 3.0 (Psychology Software Tools Inc, 2016). Participants sat behind a laptop wearing headphones and listened to pre-recorded child-directed instructions. They were told to watch the images on the screen carefully, and when they saw a moving image to click on it as fast as they could. They were also told there would be questions about the images, but it was not specified what type of questions. Then they were subjected to one of the two familiarization conditions. There was one short break after half of the familiarization trials. After the familiarization phase, participants did the test phase. They were instructed: "Look carefully at the image on the top of the screen. Which one of the two images below the white stripe looks more like the upper image?". The experimenter pointed towards the images and repeated: "Which one of these two images?". The test phase started with a practice question with filler stimuli. Participants used a computer mouse to answer the questions. Testing took approximately 10 minutes.

4.3 Results

4.3.1 Main results

Results were analysed in R (R Core Team, 2020). Practice and filler items were excluded for analysis. For every test item it was automatically recorded whether participants chose token D1 or D2 to look more like S. Overall, participants preferred token D1 over D2, but stimulus choice was influenced by familiarization condition. Figure 4.3 shows the choice for stimulus D2 per condition and includes individual variation. As the data could be conceived of as being binomially distributed, a generalized logistic linear mixed-effects model (from the package *lme4:* Bates et al., 2015) was constructed to test this finding statistically. The dependent binary variable was the choice for stimulus D2 (coded as 1, D1 was coded as 0). All eight answers for all participants were taken into account. Condition was a between-participant predictor, which was coded into sum-to-zero orthogonal contrasts (Kraemer & Blasey, 2004): -½ for Condition 2 and +½ for Condition 1. A counterbalancing predictor, PositionD2, represents the position of token D2 on the screen in the test items. This is a within-participant predictor and was coded +½ for Left and -½ for Right. The model includes by-participant random intercepts, as well as by-participant random slopes for PositionD2.

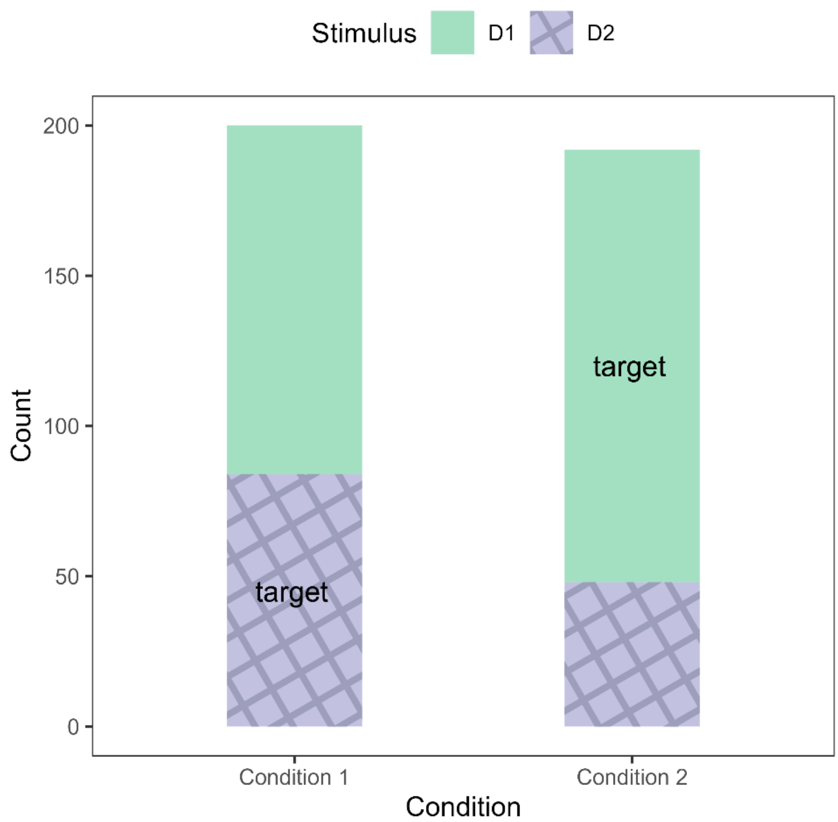


Figure 4.3 – Plot depicting the choice for stimulus D1/D2 depending on familiarization condition. "Target" indicates the target answer, which is D2 for participants in Condition 1 and D1 for participants in Condition 2. Each data point is a single trial.

We predicted that children in Condition 1 tend to categorize tokens S and D2 together while children in Condition 2 tend to categorize tokens S and D1 together. In line with this prediction, participants in Condition 1 were 3.6 (95% CI 1.3 … 11.5) times more likely (odds ratio) to choose stimulus D2 than participants in Condition 2, and this effect of Condition was significant: z = 2.384, p = 0.017. We can conclude that familiarization condition influences the preference for combining token S with token D1 or D2, indicating that the distributional properties of the input in the familiarization phase influence categorization of the stimuli (see Figure 4.4 and Table 4.1). Participants were 2.1 (95% CI 0.97 ... 5.3) times more likely to choose stimulus D2 when it was positioned left on the screen as opposed to right, but this effect of PositionD2 was not significantly different from 1 (z = 1.781, p = 0.075).

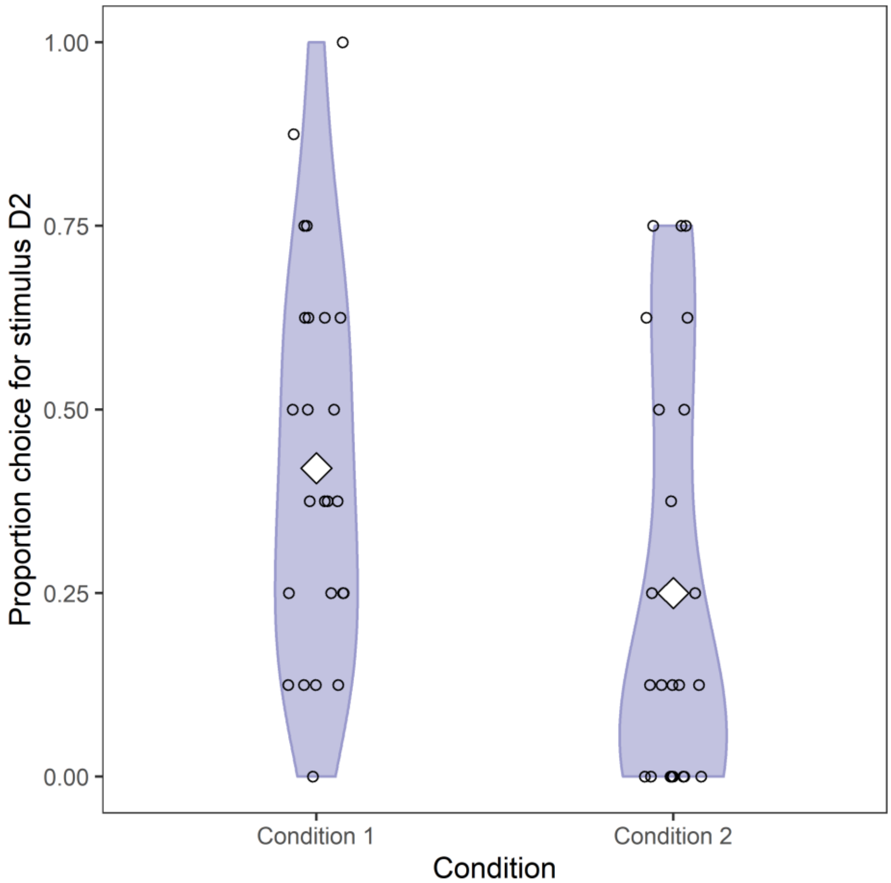


Figure 4.4 – Plot depicting the choice for stimulus D2 depending on familiarization condition which shows the individual variation.

Table 4.1 – Stimulus choice depending on familiarization condition. "Target" indicates the target answer, which is D2 for participants in Condition 1 and D1 for participants in Condition 2.

Condition	D1	D2	total
1	123	85 <u>target</u>	208
2	144 <u>target</u>	48	192

4.3.2 Results follow-up test

The design of this experiment was based on the assumption that, "a priori" (if there is no familiarization phase), stimuli D1 and D2 are equally good candidates to categorize with stimulus S. However, the results showed an overall preference for combining stimuli S and D1. As a follow-up analysis, we constructed an online experiment to test for inherent categorization preferences. A Google Form online survey was constructed, consisting of 4 questions. In every question, stimulus S was shown and participants had to choose whether they thought stimulus D1 or D2 looked more like it. The position of the two answers was counterbalanced across questions. 32 participants filled in the survey (M_{age} = 30.5 years, SD_{age} = 1.8 years)[10]. A one-sample *t*-test revealed that the probability of choosing stimulus D1 was significantly higher than chance (50%): $t = 6.506$, $p = 1.6 \cdot 10^{-9}$, (95% CI 0.67 … 0.83). This result indicates that adults have an inherent preference for categorizing stimuli S and D1 as opposed to S and D2. This could explain the unexpected overall preference for D1 in the current study, although the choice for either D1 or D2 was still significantly influenced by familiarization condition.

[10] Ideally, we would have tested new participants in the same age range as the participants of the main experiment, but due to practical constraints we tested adults. Still, the result of our follow-up test shows that the bias that we found also exists without exposure to the familiarization phase.

4.4 Discussion

In our study we aimed to investigate whether distributional learning contributes to categorizing visual stimuli in school-aged children. Familiarization condition significantly influenced categorization in our experiment leading us to conclude that children that are subjected to a familiarization phase in which certain tokens belong to one distributional peak are more likely to categorize those tokens together in the test phase as opposed to children that are trained in the familiarization condition in which other tokens belong to one distributional peak. This effect implies that children can form categories on the basis of distributional properties in the visual input.

Combining our results with those of Junge et al. (2018), we can conclude that distributional learning is important for visual object categorization in school-aged children as well as infants, at least in the absence of explicit labelling. Previous research has shown that infants, children and adults use statistical information to learn about the world. Categorization of sensory stimuli is an important process that seems to be supported by such statistical learning mechanisms. This has been shown in studies that investigate distributional learning of phonetic categories in infants, children and adults, as well as for infants learning face categories based on distributional properties. Junge et al. (2018) have shown that also novel object categories can be learned by infants based on distributional information. The present study shows that older children are sensitive to these distributional properties in the input when learning about new objects. Bottom-up statistical learning mechanisms may play a life-long role in understanding our environment.

Moreover, our study shows that the method of Chládková et al. (2022), which compared to the classic unimodal-versus-bimodal design eliminates the influence of dispersion of the tokens along the continuum, also works in the visual domain. Future studies may utilize this method to investigate (visual) distributional learning in different populations.

A small shortcoming of our study, which may have reduced its sensitivity, is the bias we found in our main experiment as well as our follow-up experiment: there seems to be an inherent preference for combining certain tokens, even without familiarization. This could be due to superficial visual properties of the stimuli. For example, the standard stimulus contains horizontal stripes, which seem to be a bit more recognizable in one distractor stimulus than the other. Interestingly, biases in arising categories are also described for children learning real-life categories (e.g. Furrer & Younger, 2005). In future studies, perhaps a different continuum of visual stimuli should be constructed and tested for inherent preferences. It might be better to choose visual stimuli that are easier to control for similarity, for example 2D shapes instead of 3D pictures. Still, the training effect remains intact, revealing that the distribution of exemplars in the familiarization phase influences novel object categorization.

Chapter 5

Distributional learning of visual object categories in children with and without developmental language disorder

This chapter is a slightly modified version of the accepted article (pending minor revisions):

Broedelet, I., Boersma, P., & Rispens, J. (2022). Distributional learning of novel visual object categories in children with and without developmental language disorder. *Submitted to Language Development Research*.

R scripts, data and materials are available on FigShare:
https://doi.org/10.21942/uva.c.5174660.v1

Abstract

It has been proposed that a deficit in statistical learning contributes to problematic language acquisition in children with developmental language disorder (DLD), but at the same time the nature and extent of this relationship is not clear. This chapter focuses on the role statistical learning in lexical-semantic development by investigating visual distributional learning of novel object categories in children with and without DLD and its relation to vocabulary knowledge. Distributional learning is a form of statistical learning and entails the learning of categories based on the frequency distribution of variants in the environment. Fifty children (25 DLD, 25 TD) were tested on a visual distributional learning task. Results indicate that children can learn novel object categories on the basis of distributional information. We did not find evidence for a deficit in visual distributional learning in children with DLD. To investigate whether visual distributional learning ability is related to vocabulary knowledge, the children with DLD were tested on different

measures of vocabulary. Phonological processing ability and non-verbal intelligence were taken into account as control variables. Multiple linear regression analyses did not reveal evidence for a relationship between distributional learning and vocabulary in DLD.

5.1 Introduction

Most children acquire their native language(s) without many major difficulties, but this is different for children with developmental language disorder (henceforth: DLD). These children do not present major neurological deficits, hearing disabilities or low overall intelligence, nor is a lack of language input the underlying problem. DLD occurs in approximately 7% of school-aged children (Bishop, 2006), and the problems often last into adulthood. Social–emotional difficulties occur in this group as well: individuals with DLD have greater risk of depression disorders (van den Bedem et al., 2019) and even have a lower quality of life compared to typically developing peers (Eadie et al., 2018).

Morphosyntactic impairments are viewed as a hallmark of DLD, while lexical abilities are often seen as a relative strength (e.g. Ullman & Pierpont, 2005). However, there is ample clinical evidence for a disadvantage in lexical skills as well (for reviews: Brackenbury & Pye, 2005; Nation, 2014). Recently, researchers have proposed that an impairment in statistical learning, a learning ability that is important for the discovery of patterns and sequences in sensory input (Siegelman, Bogaerts, Kronenfeld, et al., 2018), contributes to the language difficulties in children with DLD (Arciuli & Conway, 2018; Hsu & Bishop, 2010; Saffran, 2018). Experimental results suggest that a deficit in statistical learning (partly) explains lexical deficits (Evans et al., 2009; Mainela-Arnold & Evans, 2014), but the relationship between statistical learning and the development of lexical knowledge, especially lexical-semantic knowledge, requires more investigation. Distributional learning, which plays a role in the categorization of sensory stimuli such as speech sounds (Maye et al., 2002, 2008) and novel visual objects (Junge et al., 2018) has never been investigated in children with DLD. Categorizing novel visual stimuli might be an important skill that is required when mapping new words to new objects. In our study we aim to investigate if this type of visual distributional learning is affected in children with DLD, and whether this ability relates to different types of lexical(-semantic) knowledge.

5.2 Background

5.2.1 Statistical learning deficit hypothesis

Although the main aspect of DLD is problematic language acquisition, children with DLD experience difficulties outside the linguistic domain as well. For example, there is evidence for deficits in motor skills (Sanjeevan & Mainela-Arnold, 2019), working memory (Montgomery et al., 2010), attention (Ebert & Kohnert, 2011) and processing visual information (Collisson et al., 2015). These findings have led to the idea that a deficit in a more general learning mechanism might be at the core of the disorder, as opposed to an impairment specific to linguistic representations (Arciuli & Conway, 2018; Hsu & Bishop, 2010).

Statistical learning is such a learning mechanism that is hypothesized to be impaired in children with DLD (for a review see Siegelman, 2020). Statistical learning underlies the extraction from regularities and patterns from sensory input and has been shown to correlate with or predict language ability in children and adults (Conway et al., 2010; Ellis et al., 2014; Hamrick et al., 2018a; Kaufman et al., 2010; Kidd, 2012; Kidd & Arciuli, 2016; Misyak et al., 2010; Newman et al., 2006; Shafto et al., 2012; Spencer et al., 2015).

Results from several studies point towards a disadvantage in different types of statistical learning in individuals with DLD: learning transitional probabilities between syllables or musical tones (Evans et al., 2009; Haebig et al., 2017; Mainela-Arnold & Evans, 2014); visuo-motor sequence learning on the serial reaction time task (Lukács & Kemény, 2014; Mayor-Dubois et al., 2014; Tomblin et al., 2007); probabilistic categorization on the weather prediction task (Kemény & Lukács, 2010); implicit artificial grammar learning (Lukács & Kemény, 2014; Plante et al., 2002); non-adjacent dependency learning (Hsu et al., 2014; Lammertink et al., 2019) and visual statistical learning (Collisson et al., 2015; Lukács et al., 2021; for a review see Saffran, 2018). Please note that null results (Aguilar

& Plante, 2014; Lammertink, Boersma, Rispens, et al., 2020; Noonan, 2018) and even evidence of intact statistical learning in children with DLD (Lammertink, Boersma, Wijnen, et al., 2020) have also been reported. Importantly, several meta-analyses point to a statistical learning deficit in children with DLD in different domains (Lammertink et al., 2017; Lum et al., 2014; Lum & Conti-Ramsden, 2013; Obeid et al., 2016). Moreover, studies have suggested that statistical learning ability is related to different types of language skills in children with DLD: for example grammatical ability (Hedenius et al., 2011; Misyak et al., 2010; Tomblin et al., 2007) and lexical skills (Evans et al., 2009; Mainela-Arnold & Evans, 2014). Thus, accumulated evidence indicates that children with DLD are compromised in different types of statistical learning, which might (partly) explain their problematic language acquisition.

5.2.2 Lexical difficulties in children with DLD

Children with DLD may have difficulty with several aspects of language acquisition, such as vocabulary, morphology, syntax and phonology, and there is a large amount of heterogeneity within this population (Bishop, 2006; Leonard, 2014). Many studies have focused on morphosyntactic difficulties, for example a child saying *she walk* instead of *she walks*. However, these children also show evident difficulties in the development of lexical knowledge (Brackenbury & Pye, 2005; Nation, 2014). Research indicates that lexical difficulties impact social and academic development (Aguilar et al., 2017).

Studies suggest that children with DLD have a smaller vocabulary size and more shallow knowledge of words relative to TD children (McGregor et al., 2013). For example, they make semantic substitutions (confusing towel and blanket) and use more "all-purpose verbs" like go instead of more specific verbs like run, skip, sail, swim, etc. When naming objects, they are slower and make more phonological and semantic errors (Dockrell et al., 2001; Lahey & Edwards, 1999; Leonard et al., 1983; McGregor, 1997; McGregor et al., 2002). These errors reflect impoverished semantic representations. Dockrell et al. (2003) tested semantic knowledge of children with word-finding difficulties, and found that they provide less accurate definitions of objects and actions: their definitions often contained less information about the semantic category of an object, and more perceptual and redundant information compared to those of TD children. Moreover, compared to controls, children with DLD provide poorer, less complete definitions of common words (Mainela-Arnold et al., 2010; Marinellie & Johnson, 2002), and provide fewer semantic details in drawings (McGregor et al., 2002; McGregor & Appel, 2002).

On word association tasks, which are viewed as a measure of lexical-semantic organization, children with DLD produce fewer semantically related words than TD peers (McGregor et al., 2012; Sandgren et al., 2021; Sheng & McGregor, 2010). An inefficient lexical organization could have a negative effect on subsequent vocabulary development (Beckage et al., 2010). Finally, children with DLD also show difficulties on word learning tasks, both with learning phonological and semantic properties of words (Alt & Plante, 2006; Kan & Windsor, 2010; Nash & Donaldson, 2005) and fast mapping (Haebig et al., 2017).

Thus, children with DLD have lexical difficulties that go beyond word access, word retrieval and the phonological representations of words, pointing to suboptimal semantic representations. Little is known about the underlying cause of lexical-semantic deficits in children with DLD. Often put forward as a possible cause is poor phonological short-term memory, which is considered an important prerequisite for vocabulary acquisition (Melby-Lervåg et al., 2012). There is extensive evidence of deficits in phonological short-term memory and verbal working memory in children with DLD (for a review, see Montgomery et al., 2010). Phonological short-term memory is often measured using a non-word repetition (NWR) task. Studies show that performance on NWR tasks correlates with word-learning skills in TD children (Gathercole et al., 1997) and in children with DLD (Alt & Plante, 2006).

The causal direction of the relationship between phonological short-term memory and word learning is not clear. Difficulties with phonological processing might lead to poor phonological representations of words, which in turn may have a negative influence on the building of strong semantic representations. Indeed, NWR ability predicts vocabulary in young children between 4 and 5 years, but this relationship gets weaker in older children between 6 and 8 years (Gathercole, 2006; Gathercole et al., 1992). Furthermore, it has been found that vocabulary size is an important predictor of NWR ability, which could be explained as follows: as vocabulary size grows, phonological representations strengthen, which would improve non-word repetition ability (Metsala, 1999). Other studies fail to find evidence for a causal relationship between NWR ability and vocabulary. For example, Melby-Lervåg et al. (2012) carried out a large longitudinal study and did not find evidence for a causal relationship between NWR skills and vocabulary development in 4- to 7-year-old children. The authors also re-analysed data from a similar longitudinal study (Gathercole et al., 1992), and failed to find the causal relationship that the authors of the original study had claimed. Finally, intervention studies have failed to find an effect of phonological memory-training on vocabulary knowledge (Melby-Lervåg et al., 2012: Dahlin, Nyberg, Bäckman, & Neely, 2008, Schmiedek, Lövdén, & Lindenberger, 2010). Thus, although the difficulties in phonological processing in DLD are well-established, the role they play in vocabulary development remains unclear. Another strand of research focuses on the role of statistical learning in vocabulary development.

5.2.3 Statistical learning and the development of the lexicon

As discussed in the previous sections, a large body of studies points towards an important role for statistical learning in the acquisition of language. In children with DLD, the ability of extracting regularities from input seems to be affected, which could explain their language deficits. In this section we discuss previous research that is relevant for the relationship between statistical learning and the development of the lexicon in children with and without DLD. Specifically, we look at the link between statistical learning and lexical-semantic knowledge.

Studies have indicated that statistical learning is related to lexical development. Children with better statistical learning skills often have a larger vocabulary Spencer et al. (2015), and Shafto et al. (2012) and Ellis et al. (2014) report a predictive relationship between TD infants' performance on a visual statistical learning task and their vocabulary size at a later point in time. In another longitudinal infant study, Singh et al. (2012) found that statistical learning ability in a word segmentation task at 7 months predicts productive vocabulary at 24 months.

This relationship has also been shown for children with DLD. For example, Evans et al. (2009) reported a correlation between statistical learning ability and vocabulary knowledge in children with DLD and claimed that their lexical impairments might be explained by statistical learning difficulties. Building on this finding, Mainela-Arnold and Evans (2014) report a significant correlation between statistical learning ability on a word segmentation task and performance on a lexical-phonological access task. During this forward gating task, children heard increasingly longer parts of a word and had to guess which word they heard. On the other hand, no evidence was found for a relationship between statistical learning and performance on a word definition task. The authors suggest (from a comparison of their two p-values) that statistical learning underlies the acquisition of sequential lexical-phonological knowledge, but that lexical-semantic abilities might depend on other learning/memory systems.

Using a novel object name learning experiment, Collisson et al. (2015) showed that 3- to 4-year-old children with DLD lag behind with the development of the shape bias, which is the tendency to extend the use of newly learned object names to objects that share the same shape with the original object rather than the same colour or size. Moreover, children with DLD perform more poorly on a task that measures visual paired-associate learning, and this performance predicts the strength of their shape bias. This finding suggests that an impairment in visual statistical learning might underlie the lagging development of the shape bias in these children, in turn hindering lexical development.

In a recent study, Kemény and Lukács (2021) investigated the role of statistical learning and verbal short-term memory in passive vocabulary size in children with and without DLD. They used the weather prediction task as a measure of statistical learning. This task measures whether children can pick up a probabilistic predictive relationship between geometrical shapes and types of weather (sunshine or rain). Both groups performed above chance on the task and performance was not directly compared between groups, so the study provides no evidence for impaired statistical learning in children with DLD, in contrast to an earlier study (Kemény & Lukács, 2010). Two regression analyses were conducted with vocabulary size as the dependent variable and statistical learning, verbal short-term memory, age, receptive grammar skills, fluid intelligence as the predictor variables. For the TD children, statistical learning ability significantly contributed to vocabulary size in TD children, while verbal short-term memory did not (but please note that no direct comparison was made between the two predictors). For the children with DLD, statistical learning did not significantly contribute to vocabulary size, while verbal short-term memory did (again, no direct comparison was made between the two predictors). The authors claim that these findings suggest that lexical knowledge develops differently in children with DLD: possibly they rely more on short-term memory than statistical learning, while this is the other way around for TD children. However, it is important to note that no direct group comparison is included in the regression analysis, and thus a difference between groups cannot be claimed. This study does suggest a link between probabilistic categorization and lexical development, extending previous findings of the link between sequential word segmentation ability and vocabulary.

Considering previous work, we can conclude that the link between statistical learning and specifically lexical-semantic knowledge, with which children with DLD have been shown to have difficulties, requires further investigation. Mainela-Arnold and Evans (2014) aimed to investigate whether statistical learning contributes to lexical-phonological knowledge as well as lexical-semantic knowledge, and the authors claimed that their results show that statistical learning is related to the former but not the latter. However, the status of a potential relation cannot be concluded from comparing a null result with a statistically significant result. Moreover, a word definition task was used to measure lexical-semantic skills, which requires very explicit semantic knowledge. It could be the case that statistical learning is related to more implicit forms of semantic knowledge. Furthermore, statistical learning in this and many other studies was measured using a word segmentation task. It is not unexpected that this type of sequential statistical learning contributes to lexical-phonological knowledge due to the nature of the task. However, Mainela-Arnold and Evans (2014) also state, it is possible that other types of (non-sequential) statistical learning that were not taken into account play a role in the building of a semantically rich lexicon. Kemény & Lukács (2021) did use a non-sequential statistical learning task. However, only passive vocabulary size was measured. In the current study, we investigate whether a specific type of non-sequential statistical learning, which underlies the categorization of new visual items, is affected in children with DLD and whether it is related to different types of lexical-semantic knowledge.

While Mainela-Arnold and Evans (2014) suggested that semantic knowledge might arise from non-statistical (declarative) learning mechanisms, previous evidence suggests that statistical learning mechanisms can play a role in processing semantic information (see Paciorek & Williams, 2015 for a review). For example, the mapping of newly learned words to their corresponding referents is suggested to be a gradual statistical learning process named cross-situational learning, which entails the (implicit) tracking of co-occurrences between words and their visual referents (Kachergis et al., 2014; Smith & Yu, 2008; Suanda et al., 2014; Yu & Smith, 2011). In another strand of research, Goujon (2011) showed that adults implicitly learn that the semantic categories of real-world scenes predict the position of the following target in a visual search task, indicating that semantic information is processed automatically and can be facilitated to make unrelated decisions. Similarly, (Rogers et al., 2021) report that higher-order categories influence the learning of visual statistical regularities: people learn implicit mappings between visual stimuli better when the stimuli belonged to the same category rather than two different categories.

Another process in the development of the lexicon that could be supported by statistical learning mechanisms is learning to categorize and name the enormous number of different objects in the visual world. For example, a child needs to learn which round fruits are called apples and which ones are called peaches. Studies point out that infants automatically track the co-occurrence of visual features of objects in visual statistical learning tasks (Wu et al., 2010, 2011). This ability of learning which object features co-occur and which do not plays an important role in learning about visual categories (Palmeri & Gauthier, 2004). Similarly, Younger (1985) and Plunkett et al. (2008) showed that statistical learning may underlie semantic category learning, as infants learn object categories based on the co-occurrence of features. In the following section, we discuss distributional learning, which has been suggested to underlie categorization of phonetic and visual stimuli.

5.2.4 Distributional learning

Distributional learning is a specific type of non-sequential statistical learning which underlies categorization. Categorization is an important aspect of language acquisition in general, and forming conceptual categories and finding their label is crucial to lexical-semantic development specifically. Therefore, the current study focuses on investigating distributional learning in children with and without DLD. In this section, the principles of distributional learning will be reviewed. In a seminal study, Maye et al. (2002) showed that infants can pick up speech sound categories based on the frequency distribution of speech sound exemplars. In the original distributional learning experiment design, participants were exposed to either a unimodal or a bimodal distribution of variants from a continuum, in this case the /ta/-/da/ continuum. In the bimodal condition there were two distributional peaks, reflecting two distinct sound categories /t/ and /d/, while in the unimodal condition there was only one peak reflecting one broad category. After familiarization it was tested whether the infants could discriminate the endpoint tokens of the continuum. Maye et al. found that only their participants in the bimodal condition had statistically significantly formed two distinct categories, as they were able to discriminate the two endpoint tokens, while infants in the unimodal condition did not reach significance. This result indicated to Maye et al. that infants can learn phonetic categories based on distributional information. Although Maye et al.'s claim was based on a p-value comparison (a direct comparison between the two groups gave a non-significant p-value of 0.063), together with later findings of distributional learning of sound categories (Escudero et al., 2011; Hayes-Harb, 2007; Maye et al., 2008; Vandermosten et al., 2019; Wanrooij et al., 2014), the results point towards a distributional learning mechanism underlying bottom-up categorization of speech sounds.

More recent studies using the same experimental design have shown that distributional learning mechanisms also play a role in the visual domain, for example in categorizing new faces. In the study of Altvater-Mackensen et al. (2017), infants were subjected to a familiarization phase in either a unimodal or a bimodal condition. They saw tokens from a continuum that was created from two female faces. After familiarization, results from a discrimination task indicated that infants in the bimodal condition form two distinct categories of faces, while infants in a unimodal condition form one broad category. The same result has been shown in a novel visual object category learning experiment (Junge et al., 2018): infants in the bimodal condition showed better discrimination of two endpoint tokens than infants in the unimodal condition.

Although the results from distributional learning studies are highly interesting, it is important to note that Wanrooij et al. (2015) discuss potential pitfalls in the typical design employed when comparing a unimodal with a bimodal familiarization phase in distributional learning tasks. Namely, there might be a confounding factor at play: besides the number of distributional peaks in the input, the spreading of variants (or dispersion) also differs between conditions. This difference might result in easier discrimination of endpoint tokens for individuals who had been familiarized with the bimodal condition, as spreading of the variants is higher in that condition. To tackle the possible confounding factor, Chládková et al. (2022) designed a (auditory) distributional learning task that tackled this problem: they constructed two bimodal learning conditions which differed in the position of the distributional peaks, ensuring that spreading of the variants was not different in the two conditions.

Distributional learning thus seems to be important for the categorization of different types of sensory stimuli: speech sounds, faces and novel objects. In the current study we aim to investigate whether children with DLD have a deficit in visual distributional learning and whether this ability correlates with their lexical-semantic knowledge, as a lessened sensitivity to regularities in object categories could contribute to their problems in building strong semantic representations. Considering the possible flaw in the original design, we applied the bimodal comparison design of Chládková et al. (2022) to the visual distributional learning task of Junge et al. (2018).

5.3 The current study

The current study aims to explore the relationship between statistical learning and lexical-semantic knowledge in children with and without DLD by investigating visual distributional learning and its relation to vocabulary knowledge in children with DLD.

Our first research question was: are children with DLD less sensitive to distributional cues compared to TD children when learning novel visual object categories in an experiment? Distributional learning has never been investigated in individuals with DLD, but one study shows that distributional learning of speech sounds is impaired in children with dyslexia (Vandermosten et al., 2019). Developmental dyslexia and DLD are distinct but overlapping disorders (Snowling et al., 2020) and together with previous evidence showing that both verbal and visual statistical learning is impaired in children with DLD, we expected that they show less proficiency in visual distributional learning as well.

Our second research question was: Does the ability of visual distributional learning contribute to lexical knowledge in children with DLD? The underlying cause of the lexical-semantic difficulties in this group is not clear. There is extensive evidence for problems with phonological short-term memory, but this does not seem to be an adequate explanation. We expected that visual distributional learning contributes to these lexical-semantic difficulties, as it could be important for learning semantic information about (the use of) words, object categories and how to map words to objects. Difficulties with processing visual patterns in the environment might result in problems with building a semantically rich lexicon. Indeed, there is previous evidence for a relationship between visual statistical learning and vocabulary development (Ellis et al., 2014; Shafto et al., 2012).

To answer our research questions, we constructed a visual distributional learning task based on Junge et al. (2018) to test novel object categorization in children with and without DLD. Moreover, we measured lexical knowledge comprehensively in the children with DLD: besides productive and receptive vocabulary size, we tapped the organization of the lexicon and the knowledge of relationships between concepts/words. Finally, we control for variation in phonological processing, as children with DLD are known to have difficulties with this ability and because it is probably related to lexical knowledge. We also controlled for variation in non-verbal intelligence.

5.4 Method

5.4.1 Participants

27 children diagnosed with DLD participated in our research. One child did not finish the statistical learning task and another child was removed because of bilingualism, resulting in a final sample of 25 children with DLD (17 male, 8 female) between the ages of 7;2 and 9;3 (years;months). For the control group we used previously collected data from a study in which TD children were tested on the same task (see Chapter 4Broedelet et al., 2022)[11]. We selected 25 children (15 male, 10 female) from a larger

[11] We planned to test a new group of TD children matched to the DLD group. Unfortunately, we were unable to administer the tests as all primary schools in the Netherlands were closed from March to June 2020 due to the outbreak of COVID-19. After the reopening of the schools many restrictions still applied, making it impossible to enter schools for testing participants. We therefore decided to use a subset of an already collected dataset as control data. This dataset was previously used for an article about visual distributional learning in TD children (Broedelet et al., 2022). The decision to use previously collected data was taken only because of this circumstance, and not

sample that matched the DLD group best regarding age and gender. Their ages varied between 7;6 and 8;9. Age did not differ significantly between groups (TD age in months $M = 97.64$, $SD = 4.99$, DLD age in months $M = 96.56$, $SD = 6.49$), as tested with a two-sample t-test: $t = 1.864$, $p = 0.063$.

The children with DLD were recruited via different institutions in the Netherlands: Pento, Royal Dutch Auris Group and VierTaal. All children had been officially diagnosed with DLD by a professional clinician and were included if they met the standard DLD inclusion and exclusion criteria used within the institution. All children met the following criteria: they scored at least 1.5 standard deviations below the age norm on at least two of the four language domains (speech, auditory processing, grammar, lexical-semantic development), tested with standardized tests like the CELF; their language disorder was not secondary to a physiological or neurological disorder such as ASD, ADHD or hearing difficulties; they did not have a severe form of dyspraxia and at least one of their caretakers had acquired Dutch as a native language. Data from one child was removed because he was growing up bilingually and answered multiple questions on a vocabulary task in English.

The TD children were recruited via two primary schools in the Netherlands and met the following criteria: they had not been diagnosed with hearing difficulties, language disorders, dyslexia, ADHD or ASD and had at least one caretaker that was a native speaker of Dutch. Our study was approved by the Ethical Committee of the Faculty of Humanities of the University of Amsterdam. The parents/caretakers of all children filled in an informed consent form prior to their participation.

To get a general estimate of the language ability in our DLD subgroup, we administered the Sentence Recalling subtask from the CELF

---

because we found a significant effect in this group and deemed it sufficient to use this data. As a result of this reuse, the control group, unlike the DLD group, was not tested on the background tasks measuring vocabulary, morphosyntactic skills, phonological processing and non-verbal intelligence. This means the control group could unfortunately not be matched on vocabulary skills to the DLD group.

(Clinical Evaluation of Language Fundamentals: Core Language Scales, Dutch version; Semel et al., 2010). In this task, children are asked to repeat sentences of increasing complexity, measuring their morphosyntactic abilities. The Raven Progressive Matrices task was administered to measure non-verbal intelligence (Raven et al., 2003). One of the children could not finish the Sentence Recalling task due to time constraints. The children's scores (raw, percentile and if available norm and age-equivalent scores) on these two tasks are shown in Table 5.1. The children with DLD had low scores on the Sentence Recalling task and performed on average 50 months below their age level, confirming that our sample indeed had difficulty with language acquisition, while they scored within the average range on non-verbal intelligence. This discrepancy between language skills and non-verbal cognitive skills is typical for children with DLD.

Table 5.1 – Scores of the children with DLD on the sentence recalling and non-verbal intelligence task.

Task	Raw scores	Norm scores	Percentile scores	AES	Diff.
Sentence Recalling (N=24)	4 .. 42 *M* = 18.46 *SD* = 9.27	1 .. 8 *M* = 3.58 *SD* = 2.02	0.1 .. 25 *M* = 4.07 *SD* = 6.37	36 .. 83 *M* = 45.79 *SD* = 13.05	-68 .. -21 *M* = -50.46 *SD* = 14.43
Raven's progressive Matrices	11 .. 38 *M* = 23.24 *SD* = 7.41		5 .. 95 *M* = 41.04 *SD* = 26.25		

Notes: AES = Age-equivalent score (months). Diff. = Difference AES and chronological age. The chronological age (months) is subtracted from the age equivalent score (months). A negative value means that the age-equivalent score was lower than the actual age (M = 96.56, SD = 6.61, range 86 - 111). Scale used for interpreting percentile scores: 0-3 Very low, 3-10 Low, 10-16 Below average, 16-84 Average, 84-90 Above average, 90-98 High, 98-100 Very high. The Sentence Recalling percentile score is in the low range; the Raven's percentile score is in the average range.

5.4.2 Stimuli and design distributional learning task

The design of this experiment follows Junge et al. (2018) and Chládková et al. (2022), and was previously reported in Broedelet et al. (2022); see Chapter 4. The aim of our experiment was to measure whether the frequency distribution of tokens along a continuum influenced categorization of those tokens. To this end we constructed an 11-step continuum by morphing two pictures in equal steps using the Sqirlz 2.1 software (Xiberpic.com). We obtained permission to use the pictures of two cuddly toys from Giant Microbes (www.giantmicrobes.com) that were also used in the study of Junge et al. (2018). See Figure 5.1.

Figure 5.1 – Novel object continuum used in the experiment.

In the familiarization phase of the experiment, stimuli from the continuum were presented to the children. Two different between-participant familiarization conditions were constructed (see Figure 5.2). Not all tokens occurred equally often in the familiarization phase. The conditions differed in the exact frequency distribution of the tokens of the continuum: which token occurred more frequently compared to others. Both conditions contained a bimodal distribution with two peaks (the most frequent tokens), but the conditions differed concerning the position of the peaks in the continuum. Three of the 11 tokens, which were all equally frequent in both conditions, were used to measure categorization in the test phase: 6, 4 and 8, hereafter referred to as S (standard), D1 (deviant 1) and D2 (deviant 2).

In Condition 1 (Figure 5.2, blue line), token S and token D2 belonged to the same peak, while token 5 was shown less frequently, creating the perception of a category boundary. In Condition 2 (Figure 5.2, orange line), token S and token D1 belonged to the same peak and token 7 was shown less frequently. Our hypothesis was that our participants would learn that tokens in one distributional peak belong to one category while tokens from different peaks belong to two different categories. Therefore, we predicted that children in Condition 1 learn that tokens S and D2 belong to one category while children in Condition 2 learn that tokens S and D1 belong to one category.

Children were subjected to 12 blocks of 24 stimuli each (288 stimuli in total), as well as 2 filler stimuli per block (see Figure 5.4). In each block, the tokens of the continuum were presented one by one following the frequency distribution shown in Figure 5.2, in a randomized order. Each stimulus was shown for 800 ms and the interstimulus interval was 200 ms (based on the results of (based on Arciuli & Simpson, 2011; Turk-Browne et al., 2005). Stimuli were shown against a grey background (see Figure 5.3). A cover task was added to the task to make it more engaging: the filler stimuli jumped across the screen and children were instructed to click on them as fast as possible.

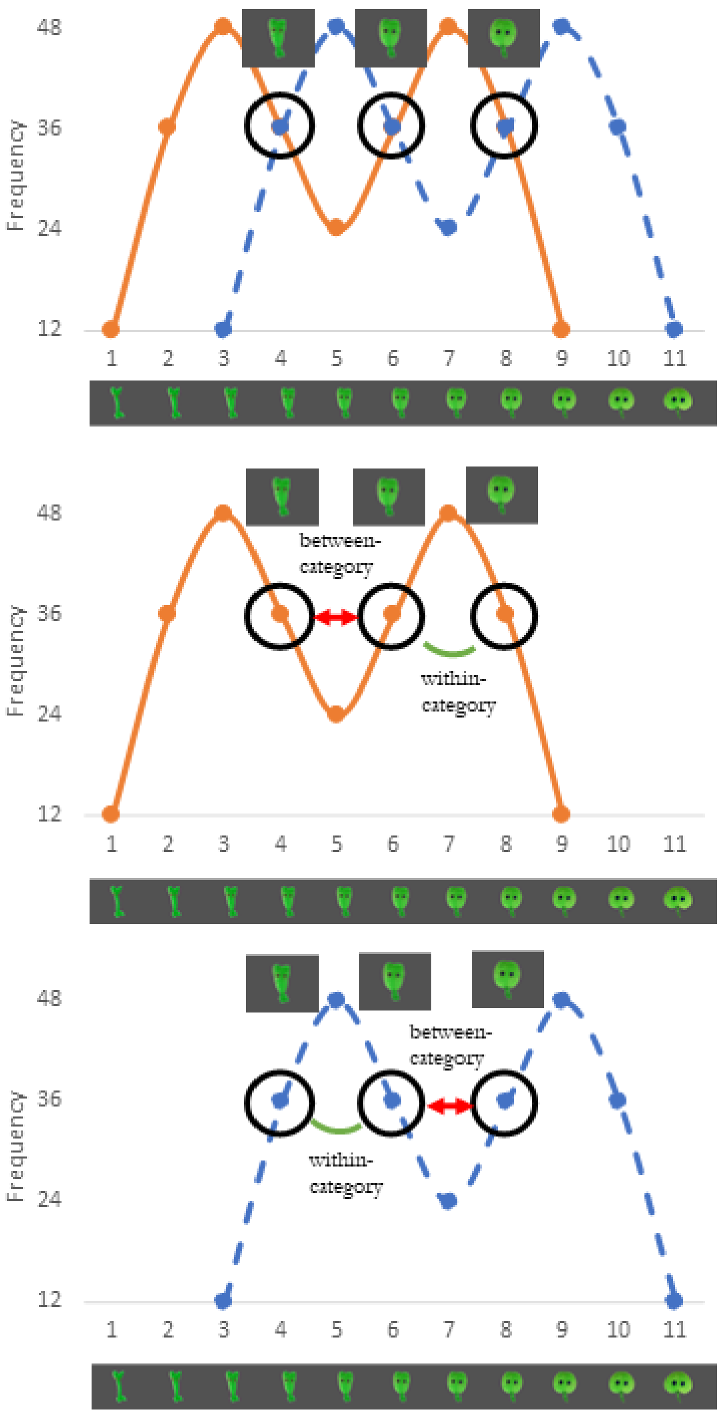

Frequency
48
36
24
12
1 2 3 4 5 6 7 8 9 10 11
Frequency
48
36
24
12
between-category
within-category
1 2 3 4 5 6 7 8 9 10 11
Frequency
48
36
24
12
between-category
within-category
1 2 3 4 5 6 7 8 9 10 11

Figure 5.2 – Familiarization conditions in the experiment. In Condition 1 (blue line), tokens S and D2 belong to one distributional peak while D1 lies in another peak. On the other hand, in Condition 2 (orange line), tokens S and D1 belong to one distributional peak while D2 lies in another peak. We hypothesize that participants in Condition 1 will learn that S and D2 belong to one category and thus will look more alike than S and D1, and the reversed for participants in Condition 2.

Figure 5.3 – A familiarization trial.

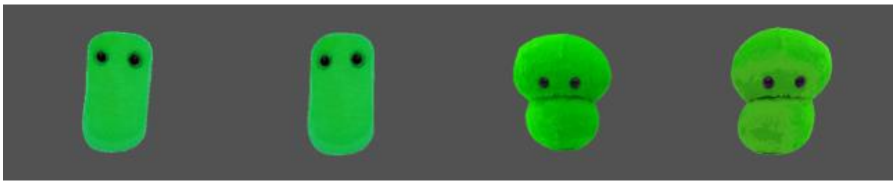


Figure 5.4 – Stimuli that were used as fillers / cover task.

Categorization was tested after familiarization using AXB-type questions. Children were asked to choose whether stimulus D1 or D2 looked more like stimulus S. In the eight questions, stimulus S was shown above a white stripe and stimuli D1 and D2 were shown below the stripe (see Figure 5.5). The position of D1 and D2 (left/right) was counterbalanced. Four filler questions were included to add some variation to the test phase, as well as a practice question. For these questions the stimuli that functioned as fillers in the familiarization phase were used and there was a clearly correct answer. The test phase was identical for every child, except that the order of the test questions was randomized. We hypothesized that children that underwent Condition 1 of the familiarization phase would choose stimulus D2 more often than children in Condition 2. This effect of Condition would be considered a learning effect.

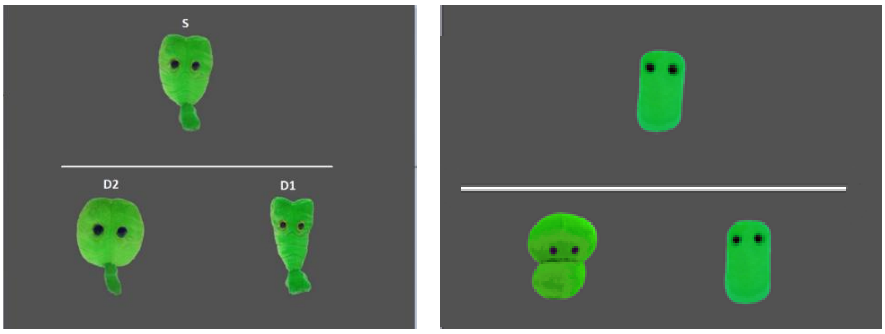


Figure 5.5 – A test question and a filler /practice question.

5.4.3 Measures of vocabulary, phonological processing, non-verbal intelligence and socio-economic status

To investigate the relationship between visual distributional learning and lexical skills in children with DLD[12], we administered several subtests of the CELF (Active Vocabulary, Word Classes 1 or 2 (depending on the age of the child) and Word Associations, as well as the Peabody Picture

---

[12] Our original plan was to investigate this relationship in both groups of children. Unfortunately, as is mentioned in our first footnote, we were not able to test the TD children on these tasks.

Vocabulary Task (PPVT; Schlichting, 2005). The tasks were used as measures of receptive and productive vocabulary size (PPVT, Active Vocabulary), the ability to find and express semantic relations between words/concepts (Word Classes 1 and 2) as well as the ability to name words of a semantic category as an indicator of lexical-semantic organization (Word Associations). See Table 5.2 for more information about the vocabulary tasks.

As control tasks, the children were tested on phonological short-term memory using the digit span task Number Repetition 1 from the CELF, on verbal working memory using the Number Repetition 2 task (digit span backwards) from the CELF and on verbal short-term memory using the non-word repetition task (Rispens & Baker, 2012). Moreover, performance on the Raven Progressive Matrices task was used as a control variable for non-verbal intelligence. See Table 5.3 for more information about the control tasks. Finally, as socio-economic status (SES) may play a role in vocabulary development (e.g. Hoff, 2003) we took the SES of the children into account using a database from Sociaal en Cultureel Planbureau (2018); no longer available. In this database, socio-economic scores are computed on the basis of the average education level and income in a particular zip code. The SES scores are based on the home addresses of the children.

Table 5.2 – Vocabulary measures administered to the children with DLD.

Construct	Task	Description	Scoring	Score range
Vocabulary size	Receptive vocabulary (PPVT)	Children heard a word and had to point to one of the four pictures.	Correct: 1 point Incorrect: 0 points	0 .. 204
Vocabulary size	Productive vocabulary (CELF)	Children saw a picture and had to name it.	2 points for a correct answer, for some items there were 1-point answer possibilities	0 .. 56
Semantic knowledge	Word Classes 1 (7 y.o. children) (CELF)	Children had to choose which two out of three/four pictures were related and why.	1 point for choosing the correct picture, 1 point for expressing the relationship correctly	0 .. 38
Semantic knowledge	Word Classes 2 (8+) (CELF)	Children had to choose which two words out of four were related and why.	1 point for choosing the correct word, 1 point for expressing the relationship correctly	0 .. 40
Lexical-semantic organization	Word Associations (CELF)	Children had to name as many words as they could in a semantic category: food, clothes and professions.	1 point for every related word	0 .. ∞

Table 5.3 – Control measures administered to the children with DLD.

Construct	Task	Description	Scoring	Score range
Verbal short-term memory	Digit span forwards	Children had to repeat strings of number increasing in length.	Correct: 1 point Incorrect: 0 points	0 .. 16
Verbal working memory	Digit span backwards	Children had to repeat strings of number backwards increasing in length.	Correct: 1 point Incorrect: 0 points	0 .. 14
Phonological short-term memory	Non-word repetition	Children had to repeat non-words.	Correct: 1 point Incorrect: 0 points	0 .. 22
Non-verbal intelligence	Raven Progressive Matrices	Children had to complete a visual pattern.	Correct: 1 point Incorrect: 0 points	0 .. 60

5.4.4 Procedure

Testing took place in a quiet room in the school or in the home of the child. The distributional learning experiment was run on a laptop computer using E-Prime 3.0 (Psychology Software Tools Inc, 2016). Children wore headphones. We had recorded the instructions in advance, in a child-directed manner. Before the experiment started, the children were instructed to look at the images on the screen and click on moving images as fast as they could if they saw one. They were told to watch carefully as there would be questions about the images later on, but the type of questions was not specified. The experiment started when the child confirmed that s/he understood the task. Familiarization condition was counterbalanced between participants. There was a short break halfway the familiarization phase and the child could indicate when s/he wanted to continue. The test phase started immediately after the familiarization phase with a practice question. Children were instructed to carefully look at the image above the white stripe, and to indicate which of two images below the stripe they thought looked more like the upper image. The experimenter repeated the question while pointing out the images. The experiment had a total duration of approximately 10 minutes.

Besides the distributional learning task, the children with DLD did two other statistical learning tasks (results of one of those tasks are discussed in Chapter 3) as well as the aforementioned background tasks. For those children, testing was divided over two separate test sessions on different days; the second session usually took place within a few days or one week. The order of the tasks within the sessions as well as the order of the sessions was counterbalanced across participants. Each test session took approximately 50 to 60 minutes.

5.5 Results

5.5.1 Split-half reliability distributional learning task

Split-half reliability was computed as a measure of reliability of the distributional learning task. Two separate generalized mixed effect models were run with only the odd or even test items included. Then, the correlation between the answers to even and odd test items was computed, using the random slopes of the intercept for the even/odd test items. After the application of the Spearman-Brown correction, the split-half reliability of the task turned out to be r = 0.73 (95% CI 0.52 .. 0.85), approaching the value of r = 0.80 which is considered the standard that reliable tests should meet (Nunnally & Bernstein, 1994).

5.5.2 Group comparison distributional learning task[13]

See Table 5.4 and Figure 5.6 for the descriptive data. As a first step in our analysis, we removed all practice and filler items from the data. A generalized mixed effect model was run with the package *lme4* (Bates et al., 2015) in R (R Core Team, 2020) to test whether familiarization condition and participant group influenced categorization. The choice for stimulus D2 (which could either be 1 or 0) was the dependent variable. Between-participant predictors were Condition (Condition 1 or 2), Group (TD or DLD) and Age (in months). PositionD2 was a within-participant predictor reflecting the position of token D2 (left or right) that varied

[13] *The TD children of whom results are reported here are a subgroup of the sample reported in* Broedelet et al. (2022).

between test items. We chose the maximal model that is still correctly computable and that keeps all its included predictors and interactions reportable (by including random slopes for all within-participant predictors and interactions). The model includes main effects for Condition, Group, Age and PositionD2, all two- and three-way interactions between Condition, Group and Age as well as the simple interaction between Condition and PositionD2. Moreover, we included random intercepts by participant as well as by-subject random slopes for PositionD2. Sum-to-zero orthogonal coding (Kraemer & Blasey, 2004) was applied to the predictors Condition (-½ for Condition 2 and +½ for Condition 1), Group (-½ for DLD and +½ for TD) and Position D2 (-½ for right and +½ for left). The predictor Age was centred by subtracting its average.

Table 5.4 – Descriptive data for the choice of stimulus D1 or D2.

	TD children		Children with DLD	
	D1	D2	D1	D2
Condition 1	55	49 <u>target</u>	61	35 <u>target</u>
Condition 2	71 <u>target</u>	25	84 <u>target</u>	20

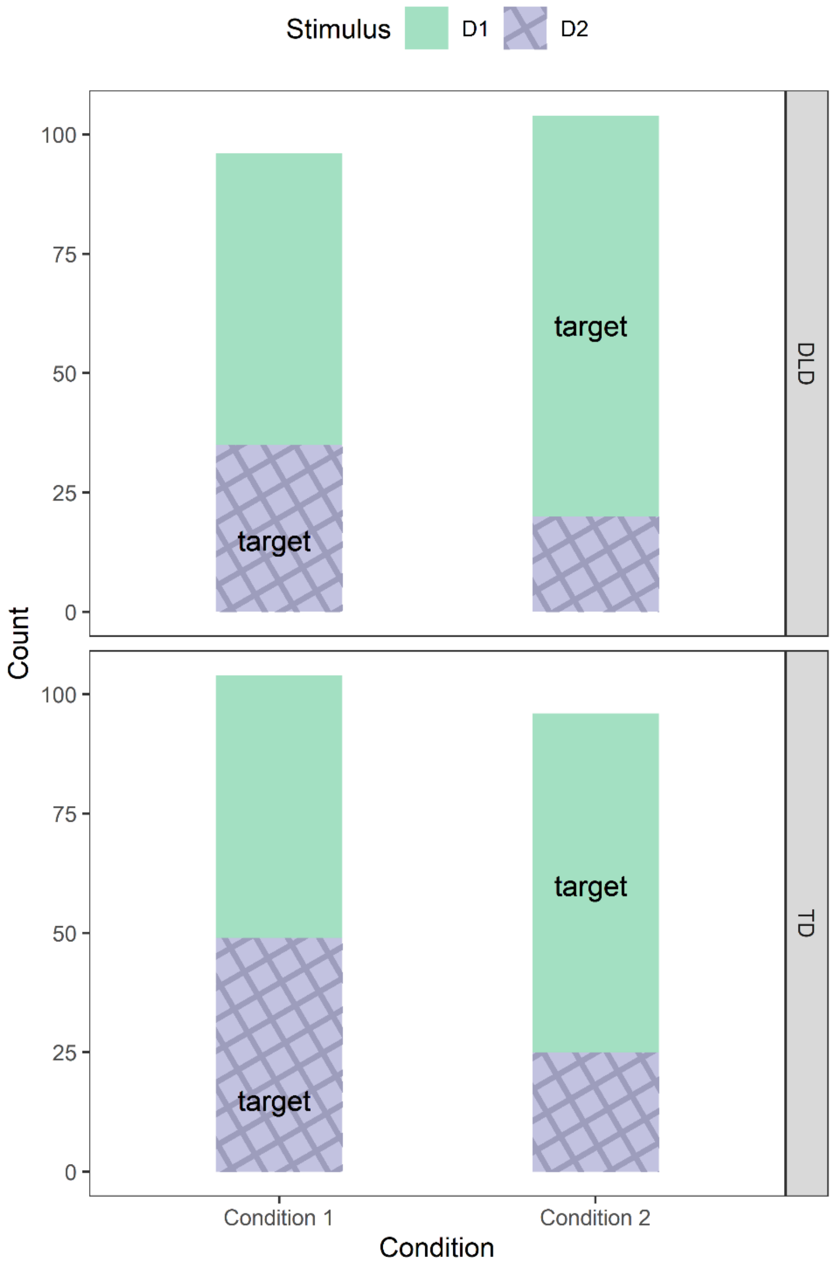


Figure 5.6 – Choice for stimulus D1 / D2 per Group and Condition.

We predicted that if children are sensitive to the distributional cues in the familiarization phase, our children in Condition 1 would prefer the combination S + D2, while our children in Condition 2 would prefer the combination S + D1 - in other words, a stronger preference for D2 in Condition 1 than Condition 2. This could manifest as a significant effect of Condition on the dependent variable. Moreover, we expected that our children with DLD would be less sensitive to the distributional cues in the familiarization phase than our TD children, which would manifest as a significant interaction between the effects of Condition and Group on the dependent variable, indicating that the Condition effect is not equally strong in the two subpopulations.

Confirmatory results. In our sample, as determined by our model, Condition influenced the choice for stimulus D2: children in Condition 1 were 4.04 times more likely to choose stimulus D2 than children in Condition 2, and this effect was significantly above 1: $z = 2.758$, $p = 0.006$, 95% CI 1.497 .. 10.9. This is in line with our prediction and indicates that school-aged children can learn novel visual object categories based on distributional properties. Our second prediction is not confirmed: although the effect of Condition was 1.007 times stronger in the TD group compared to the DLD group, this interaction between Condition and Group was not significantly above 1: $z = 0.007$, $p = 0.994$, 95% CI 0.15 .. 6.8. We thus cannot conclude anything about a difference in distributional learning in children with DLD compared to TD children: the confidence interval tells us that children with DLD could be up to 6.7 times better or 6.8 times weaker on the visual distributional learning task than TD children. We therefore cannot conclude whether children with DLD do or do not have a distributional learning deficit. See Figure 5.7 for a plot depicting the choice for stimulus D2 depending on Condition and Group.

Exploratory results. To explore whether children with DLD show learning on the visual distributional learning task, we ran a separate model which only included the children with DLD. This model included the main effects for Condition, Age and PositionD2 as well as all three-way interactions between those predictors. According to the model, our children with DLD in Condition 1 were 3.75 times more likely to choose D2 than our children in Condition 2, but the effect was not significantly above 1: $z = 1.788$, $p = 0.074$, 95% CI 0.86 .. 19.4[14]. On the basis of this result, we cannot conclude whether children with DLD are able to learn novel visual object categories based on distributional information.

[14] When we ran a model which included random slopes per participant for PositionD2 (as we did in our first model with all participants), the effect of Condition was 4.11 (95% CI 1.01 .. 16.7): $z = 1.977$, $p = 0.048$. However, as this model had a singular fit, we chose to report the results of a simplified model without random slopes for PositionD2 (this makes the effect of PositionD2 unreportable, but as we are not directly interested in this effect, this is not problematic). Note that neither the *p*-value of 0.074 neither the *p*-value of 0.048 can be called statistically significant, because this exploratory test came on top of the earlier confirmatory test, for which we already used a preset *p*-value criterion of 0.05.

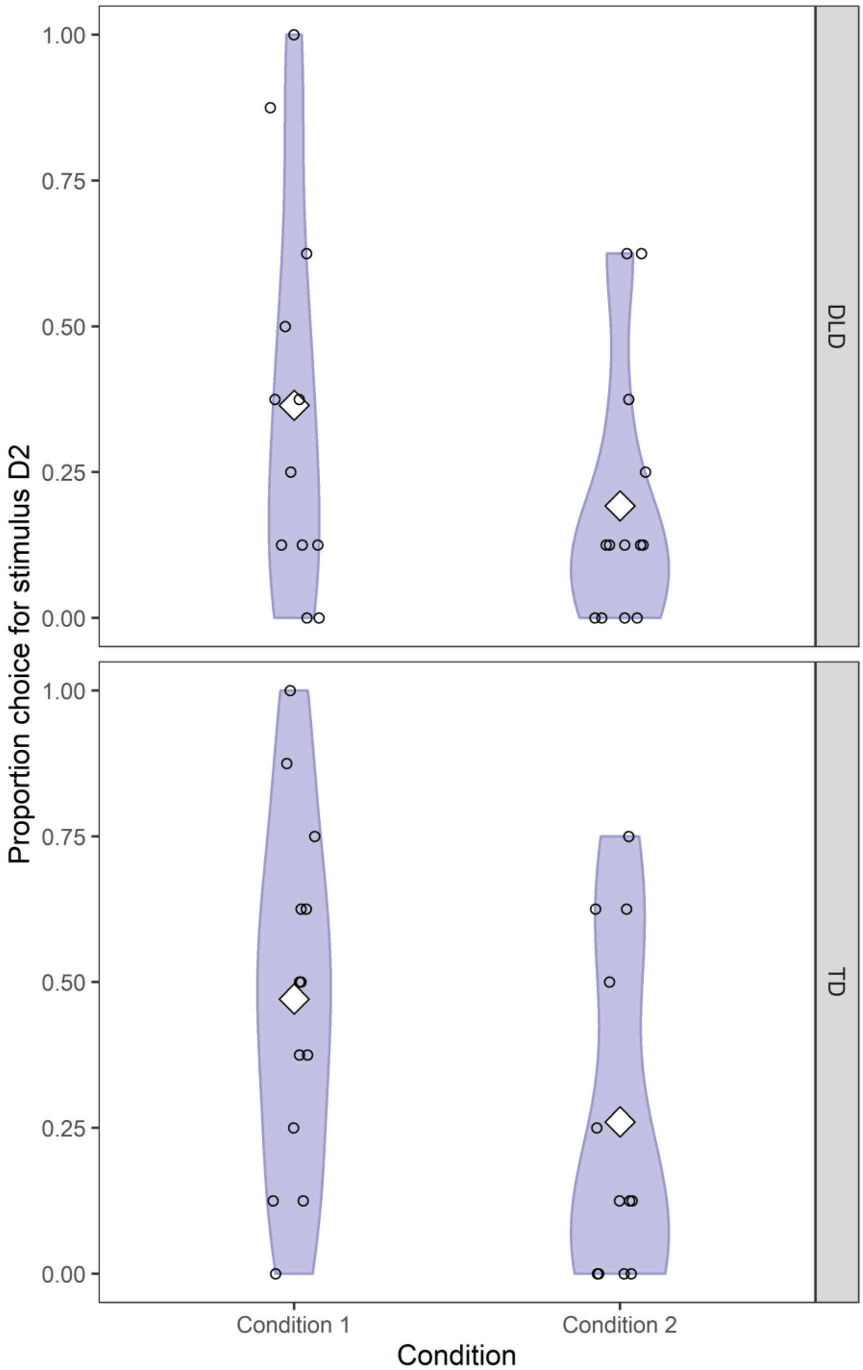


Figure 5.7 – Choice for stimulus D2 per Group and Condition.

5.5.3 Regression analyses

Descriptive data. To investigate the relationship between distributional learning and vocabulary, we administered tasks measuring several types of lexical knowledge to the children with DLD, as well as several control tasks (see section 4.3). In Table 5.5 we present the scores of the children with DLD on the vocabulary tasks and in Table 5.6 their scores on the control tasks: the raw scores, the norm and percentile scores (if available), and the age-equivalent scores. The raw scores are used in our statistical analysis. The norm, percentile and age-equivalent scores are presented to illustrate the abilities of the children with DLD.

Table 5.5 – Children with DLD's scores on the vocabulary tasks

Task	Subtask	Raw scores	Norm scores	Percentiles	AES	Diff.
Productive vocabulary		8 .. 41 *M* = 28.16 *SD* = 8.94	2 .. 12 *M* = 6.84 *SD* = 2.46	0.4 .. 75 *M* = 20.46 *SD* = 21.33	36 .. 98 *M* = 73.24 *SD* = 16.69	-62 .. 7 *M* = -23.32 *SD* = 15.83
Receptive vocabulary		70 .. 119 *M* = 90.48 *SD* = 13		0 .. 91 *M* = 27.36 *SD* = 26.82		
Word associations		10 .. 42 *M* = 23.92 *SD* = 6.37	2 .. 15 *M* = 7.48 *SD* = 2.45	0.4 .. 95 *M* = 24.22 *SD* = 19.77	42 .. 133 *M* = 77.2 *SD* = 18.42	-56 .. 42 *M* = -19.36 *SD* = 19.19
Word classes	Receptive	2 .. 19 *M* = 11.2 *SD* = 6.95	3 .. 12 *M* = 7.24 *SD* = 2.63	1 .. 75 *M* = 24.92 *SD* = 22.6	36 .. 109 *M* = 70.36 *SD* = 22.85	-68 .. 18 *M* = -26.2 *SD* = 25.59
	Expressive	0 .. 18 *M* = 8.8 *SD* = 5.95	1 .. 13 *M* = 6.88 *SD* = 2.71	0.1 .. 84 *M* = 21.68 *SD* = 22.84	36 .. 116 *M* = 71.92 *SD* = 19.19	-68 .. 25 *M* = -24.64 *SD* = 21.06
	Total	2 .. 37 *M* = 20 *SD* = 12.79	2 .. 13 *M* = 6.88 *SD* = 2.60	0 .. 84 *M* = 21.3 *SD* = 22.57	36 .. 116 *M* = 71.6 *SD* = 19.4	-68 .. 25 *M* = -24.96 *SD* = 21.72

Notes: AES = Age-equivalent score (months). Diff. = Difference AES and chronological age. The chronological age (months) is subtracted from the age equivalent score (months). A negative value means that the age-equivalent score was lower than the actual age (M = 96.56, SD = 6.61, range 86 - 111). Scale used for interpreting percentile scores: 0-3 Very low, 3-10 Low, 10-16 Below average, 16-84 Average, 84-90 Above average, 90-98 High, 98-100 Very high. The scores for the vocabulary tasks fall within the average range.

Table 5.6 – Children with DLD's scores on the control tasks.

Task	Subtask	Raw scores	Norm scores	Percentile scores	AES	Diff.
Raven's progressive Matrices		11 .. 38 *M* = 23.24 *SD* = 7.41		5 .. 95 *M* = 41.04 *SD* = 26.25		
Digit Span	Forwards	3 .. 9 *M* = 5.36 *SD* = 1.58	1 .. 12 *M* = 6 *SD* = 2.8	0.1 .. 75 *M* = 16.6 *SD* = 21.27	50 .. 103 *M* = 68.76 *SD* = 16.87	-52 .. 12 *M* = -27.8 *SD* = 17.33
	Backwards	0 .. 4 *M* = 2.72 *SD* = 1.02	2 .. 11 *M* = 7.52 *SD* = 2.35	0.4 .. 63 *M* = 26.06 *SD* = 19.34	57 .. 101 *M* = 79.52 *SD* = 13.97	-43 .. 14 *M* = -17.04 *SD* = 13.73
	Total	4 .. 12 *M* = 8.08 *SD* = 1.91	1 .. 10 *M* = 5.68 *SD* =2.39	0.1 .. 50 *M* = 12.9 *SD* = 14.67	48 .. 102 *M* = 71.8 *SD* = 11.81	-56 .. -2 *M* = -24.76 *SD* = 12.31
Non-word repetition		0 .. 9 *M* = 3.36 *SD* = 2.36		Low		

Notes: AES = Age-equivalent score (months). Diff. = Difference AES and chronological age. The chronological age (months) is subtracted from the age equivalent score (months). A negative value means that the age-equivalent score was lower than the actual age (M = 96.56, SD = 6.61, range 86 - 111). Scale used for interpreting percentile scores: 0-3 Very low, 3-10 Low, 10-16 Below average, 16-84 Average, 84-90 Above average, 90-98 High, 98-100 Very high. The scores for the Raven, and digit span backwards fall within the average range, the scores for digit span forwards and total digit span score fall in the below average range.

In contrast to their scores on the sentence recall task (see Table 5.1), the children with DLD scored within the average range (low end of the continuum) on the measures of vocabulary. The age-equivalent scores on these subtasks were between 19.36 and 26.2 months below their chronological age. Their non-verbal intelligence scores are also within the average range (see Table 5.6). However, the children showed below-average scores on the digit span forward task, which presumably reflect limitations in phonological short-term memory, which are reported often in DLD (Montgomery et al., 2010). Norm scores are available for the non-word repetition task for TD children of 7 (N = 96) years old, 8 years old (N = 82) and 9 years old (N = 208)[15]. The mean raw scores for these age groups are 8.03, 8.83 and 9.07 out of 22 words correct respectively. Compared to that, the average score of 3.36 out of 22 in our group of children with DLD (see Table 5.6) can be considered as low. The

[15] https://progracy.com/normscores/

children's age in months was on average 96.56 (*SD* = 6.61, range 86 .. 111), and their SES score on average -0.37 (*SD* = 1.04, range = -1.96 .. 1.52).

Principle component analysis. Prior to the regression analysis, all variables were centred around zero and scaled to a standard deviation of 1. To reduce the number of predictor variables, we ran a principal component analysis (PCA) in R using the raw scores on the digit span forward, digit span backward, non-word repetition and non-verbal intelligence tasks. The PCA analysis yielded four components, which explained 44%, 36%, 15% and 5% of the variance respectively. On the basis of this outcome, we decided to use three components, as they together explained 95% of the variance in the data. After varimax rotation, the three components explained 46%, 27% and 26% of the variance respectively. These components were saved and used for further analysis. See *Table 5.7* for the component loadings. The first component represented phonological processing (mainly digit span forward and non-word repetition scores), the second component non-verbal intelligence (mainly Raven scores), and the third component verbal working memory (mainly digit span backward scores).

Table 5.7 – Standardized loadings of the varimax-rotated PCA.

	Component 1 (phonological processing)	Component 2 (non-verbal intelligence)	Component 3 (verbal working memory)
Digit span forwards	0.93	-0.22	0.05
Digits span backwards	0.05	0.20	0.98
Non-word repetition	0.95	0.13	0.03
Non-verbal intelligence	-0.05	0.97	0.21

Predictor variables. The predictor variables were accuracy on the distributional learning task, age, SES, and the three component scores representing phonological processing, non-verbal intelligence and verbal working memory respectively. There were no significant correlations between the predictor variables (see Table 5.8). Accuracy on the distributional learning task was used as the measure for distributional learning ability, and was computed by comparing the answer to every test question to the target answer. For Condition 1, the target answer was D2 while it was D1 for Condition 2. This variable thus reflects sensitivity to the distributional properties in the familiarization phase. See Figure 5.8 for the distribution of the accuracy scores.

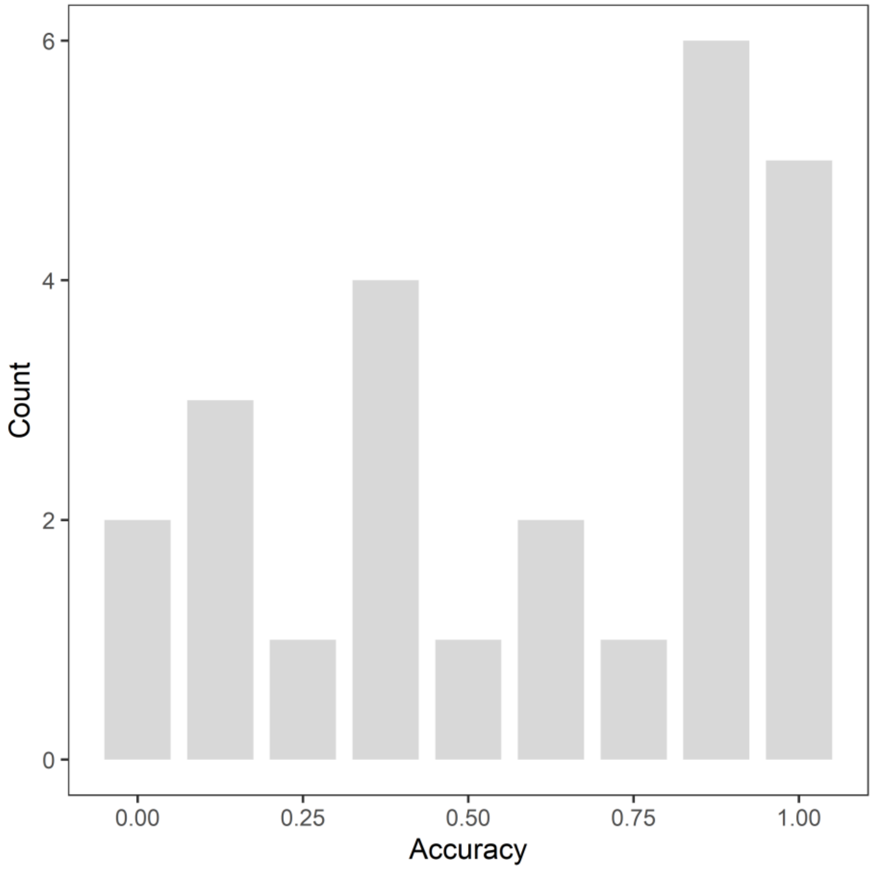


Figure 5.8 – Distribution of accuracy scores on the distributional learning task.

Table 5.8 – Correlations between the predictor variables.

	Comp 1 (phonological processing)	Comp 2 (non-verbal intelligence)	Comp 3 (verbal working memory)	Age	SES
Distributional learning	*r* = -0.17 *p* = 0.426	*r* = -0.05 *p* = 0.819	*r* = -0.24 *p* = 0.256	*r* = 0.09 *p* = 0.677	*r* = 0.03 *p* = 0.881
Component 1 (phonological processing)		*r* = 0 *p* = 1	*r* = 0 *p* = 1	*r* = 0.09 *p* = 0.675	*r* = 0.02 *p* = 0.917
Component 2 (non-verbal intelligence)			*r* = 0 *p* = 1	*r* = -0.21 *p* = 0.323	*r* = 0.02 *p* = 0.923
Component 3 (verbal working memory)				*r* = 0.26 *p* = 0.217	*r* = -0.22 *p* = 0.299
Age					*r* = 0.04 *p* = 0.866

Dependent variables. We ran four separate multiple linear regression analyses in R to test the relationship between distributional learning and different measures of vocabulary. The dependent measures were raw scores on the tasks measuring receptive vocabulary size, productive vocabulary size, and word associations. For the scores on the word classes tasks (part 1 and part 2) we decided to use the norm total scores (receptive + expressive) instead of raw scores (see Table 5.5), as the range of scores differed a lot between the children who did part 1 and the children who did part 2 of the task, which depended on their age.

The linear models. The first model was run with receptive vocabulary size as the dependent variable and the five predictors as predictor variables. The model did not explain variation in receptive vocabulary size better than the null model (F = 0.59, p = 0.734) and none of the predictors were significant (see Table 5.9). The second model with productive vocabulary size as the dependent variable also was not significant (F = 1.693, p = 0.18) and contained no significant predictors (see Table 5.10). The third model with word classes total score as the dependent variable was not significant (F = 1.604, p = 0.2033), but component 2 (non-verbal intelligence) significantly predicted word classes score (t = 2.156, p =

0.045), indicating that the ability of completing non-verbal patterns might explain unique variance in semantic knowledge about words, but please note that this result is exploratory. None of the other predictors were significant (see Table 5.11). The last model with word association score as the dependent variable was not significant (F = 0.827, p = 0.564), and none of the variables significantly predicted the dependent variable (see Table 5.12). In none of the models distributional learning significantly predicted vocabulary scores. Based on this null result, we cannot conclude anything about the relationship between visual distributional learning and vocabulary knowledge.

Table 5.9 – Results from the first linear model predicting receptive vocabulary size.

Predictor	Estimate (log odds) [95% CI]	Std. error (log odds)	*t*	*p*
Age	0.18 [-0.31 .. 0.67]	0.234	0.782	0.444
SES	0.09 [-0.37 – 0.56]	0.222	0.420	0.680
Component 1 (phonological processing)	0.27 [-0.20 .. 0.73]	0.220	1.204	0.244
Component 2 (non-verbal intelligence)	0.17 [-0.30 .. 0.63]	0.221	0.747	0.465
Component 3 (verbal working memory)	0.12 [-0.38 .. 0.62]	0.239	0.501	0.622
Distributional learning	0.15 [-0.33 .. 0.63]	0.229	0.674	0.509

Table 5.10 – Results from the second linear model predicting productive vocabulary size.

Predictor	Estimate (log odds) [95% CI]	Std. error (log odds)	*t*	*p*
Age	0.32 [-0.11 .. 0.75]	0.205	1.584	0.131
SES	0.36 [-0.05 .. 0.77]	0.194	1.869	0.078
Component 1 (phonological processing)	0.29 [-0.11 .. 0.70]	0.193	1.520	0.146
Component 2 (non-verbal intelligence)	0.14 [-0.27 .. 0.55]	0.193	0.729	0.476
Component 3 (verbal working memory)	-0.03 [-0.47 .. 0.41]	0.209	-0.126	0.901
Distributional learning	0.08 [-0.34 .. 0.50]	0.200	0.401	0.693

Table 5.11 – Results from the third linear model predicting word classes total score.

Predictor	Estimate (log odds) [95% CI]	Std. error (log odds)	*t*	*p*
Age	-0.19 [-0.63 .. 0.24]	0.207	-0.930	0.365
SES	0.04 [-0.38 .. 0.45]	0.196	0.180	0.8595
Component 1 (phonological processing)	-0.25 [-0.66 .. 0.16]	0.195	-1.301	0.2098
Component 2 (non-verbal intelligence)	0.42 [0.01 .. 0.83]	0.195	2.156	0.045*
Component 3 (verbal working memory)	0.03 [-0.42 .. 0.47]	0.211	0.124	0.903
Distributional learning	-0.17 [-0.60 .. 0.25]	0.202	-0.859	0.402

Table 5.12 – Results from the fourth linear model predicting word association score.

Predictor	Estimate (log odds) [95% CI]	Std. error (log odds)	*t*	*p*
Age	0.14 [-0.33 .. 0.62]	0.227	0.630	0.536
SES	0.23 [-0.23 .. 0.68]	0.215	1.049	0.308
Component 1 (phonological processing)	0.10 [-0.34 .. 0.55]	0.214	0.486	0.633
Component 2 (non-verbal intelligence)	0.13 [-0.32 .. 0.57]	0.214	0.584	0.567
Component 3 (verbal working memory)	-0.30 [-0.79 .. 0.18]	0.232	-1.307	0.208
Distributional learning	0.88 [-0.38 .. 0.55]	0.222	0.398	0.695

5.6 Discussion

In the current study we aimed to shed more light on the relationship between statistical learning ability and lexical-semantic skills in children with and without DLD. Specifically, we investigated whether children with DLD are sensitive to distributional information in a visual distributional learning task, and whether this ability is related to different types of lexical-semantic knowledge. Our results show that, overall, school-aged children learn novel visual object categories based on distributional information. We cannot answer our first research question as we did not find evidence for or against a visual distributional learning deficit in children with DLD. The confidence interval of our group comparison shows that children with DLD could be between 6.8 times weaker and 6.7 times better on the visual distributional learning task than TD children. The finding of a non-significant group difference could be due to chance. It is possible that the true effect is zero, but we can only speculate about possible underlying reasons.

It could be the case that children with DLD have no disadvantage in visual distributional learning compared to TD children. Previous evidence has suggested that visuomotor statistical learning is impaired in children with DLD (Lum et al., 2014; Obeid et al., 2016; Tomblin et al., 2007). Please note that null results have been found (Aguilar & Plante, 2014; Noonan, 2018) and Lammertink, Boersma, Rispens, et al. (2020) report evidence for visual statistical learning in children with DLD. Intact visual statistical learning cannot be concluded from our null result, but accumulated evidence could point towards a specifically verbal statistical learning deficit in children with DLD, as opposed to a domain-general deficit. Statistical learning is often characterized as a domain-general ability, but research suggests the existence of different domain-specific components of statistical learning (Siegelman, 2020). It is also possible that sequential statistical learning as is tested with for example word segmentation tasks is problematic for children with DLD, while specifically distributional learning is not. More research is necessary to disentangle these possibilities. For example, it would be interesting to investigate whether verbal distributional learning is problematic for children with DLD.

The absence of a significant DLD–TD difference could also be due to a lack of statistical power. We tested 25 children in both participant groups, but the between-participants design of our experiment results in relatively limited number of participants per subgroup. Future studies should test larger participant groups and/or change the design such that multiple between-participant comparisons are avoided. Another option would be to test categorization in a way that would provide more data, for example by using an on-line behavioural measure or a neurological measure like EEG (Altvater-Mackensen et al., 2017), which could make the task more sensitive to potential DLD–TD differences.

To answer our second research question, we investigated whether distributional learning ability predicted vocabulary knowledge in children with DLD, while controlling for variation in phonological processing, verbal working memory, non-verbal intelligence, SES and age. We did not find any evidence for or against this relationship in our sample of children with DLD. Apart from chance, several factors could underlie this null-result. It could be the case that, as statistical learning tasks are designed to measure group-level performance, they are not suitable for measuring individual differences reliably and thus should not be used to predict differences in language outcome (Arnon, 2019; Siegelman, Bogaerts, & Frost, 2017; Siegelman, Bogaerts, Christiansen, et al., 2017). For example, Arnon (2019) showed that three different statistical learning tasks had a low test-retest reliability and internal consistency in children, illustrating that they did not capture individual statistical learning ability reliably. This is a serious problem in the field of statistical learning research, as correlations between statistical learning ability and language proficiency might have been both overestimated and underestimated in previously reported studies (Siegelman, 2020). The split-half reliability of our visual distributional learning task was $r = 0.73$, approaching the standard of $r = 0.80$. This suggests that the test is a fairly reliable test of categorization. However, test-retest reliability should still be investigated to find out whether this task is able to capture individual differences reliably.

Another phenomenon that could occur when investigating individual differences in statistical learning is a large portion of the participants performing around chance level. Variation around chance level is not meaningful variation, which could result in the absence of significant correlations. However, this does not seem to be the case for our sample (see Figure 5.8). Another problem with this type of tasks might be that implicit knowledge that is built during familiarization does not transfer to the more explicit test questions in the test phase. Introducing more implicit and/or on-line measures of statistical learning could address this problem.

Importantly, although we did not compare the children with DLD to TD children on measures of vocabulary directly, it is striking that the percentile scores of the children with DLD in our sample are within the average range. Still, it is important to note that the ranges are wide and the children do fall behind same-aged peers if we consider the age-equivalent scores. The scores on the task measuring syntax and morphology do fall in the low range. This could mean that grammatical difficulties are more pronounced than vocabulary problems in our sample. Future studies could consider picking specific subgroups of children with DLD who have pronounced vocabulary problems to investigate the relationship between statistical learning and vocabulary development.

Although we cannot conclude this on basis of our results, there is also the possibility that there is no (strong) relationship between statistical learning and lexical-semantic knowledge. Perhaps statistical learning does contribute to more structural linguistic knowledge such as rules and regularities, but deeper (semantic) knowledge is subject to other types of learning mechanisms, although research did point out that statistical learning mechanisms are sensitive to semantic information (Goujon, 2011; Paciorek & Williams, 2015). Possibly, deficits in other cognitive mechanisms such as attention, inhibition or verbal short-term memory play a role in the lexical-semantic difficulties that are observed in children with DLD (Alt & Plante, 2006; Mainela-Arnold & Evans, 2014). More research into these difficulties and their underlying mechanisms is necessary.

We included measures of phonological processing, verbal working memory and non-verbal intelligence in our regression models as control variables. Somewhat unexpectedly, we did not find evidence for a contribution of phonological processing or verbal working memory ability to different types of vocabulary knowledge in our sample of children with DLD. Similarly, Rispens and Baker (2012) found no evidence for a relationship between non-word repetition and vocabulary size in TD children and children with DLD, and the longitudinal study of Melby-Lervåg et al. (2012) yielded no evidence of a causal relationship between non-word repetition and vocabulary acquisition in 4- to 7-year-old TD children. A meta-analysis could shed light on the relationship between phonological processing and vocabulary development in children with and without DLD. Moreover, we found an indication that non-verbal intelligence contributes to word category knowledge in children with DLD. This might be explained by similarities between the tasks: in the Word Category task, children had to choose which two out of three pictures/words were related (and why), while in the Raven progressive matrices task children had to complete visual patterns (see Table 5.3). Still, it is an interesting finding that non-verbal intelligence could explain variation in the verbal (semantic) domain, although we want to emphasize again that this is an exploratory finding.

A shortcoming of the visual distributional learning task we have used is the finding that children overall prefer the combination S + D1, which is a result we have also reported in Chapter 4. In that study, we tested 32 adults in an online experiment to explore a priori preferences for either S+D1 or S+D2. We wanted to investigate how participants who had not been exposed to a familiarization phase would answer questions similar to the test phase of our experiment. Results showed that participants chose D1 to look more like S 75% of the time, which was significantly higher than chance level. This result implies that D1 looks more like S for most participants, which is not an ideal starting point for testing the influence of distributional learning on categorization. This a priori preference might have diminished the distributional learning effect as well as a potential group difference in learning. However, our results show that despite this preference for the combination of S+D1, exposure to a familiarization phase in which S and D2 belonged to one distributional peak still caused participants to categorize S and D2 more often. Future studies might choose to use different stimuli when testing visual distributional learning and test beforehand whether participants show any unexpected preferences.

5.7 Conclusions and further directions

Our study shows that school-aged children can learn novel visual object categories based on distributional information. We did not find evidence for or against a visual distributional learning deficit in children with DLD. Future research could use our results for meta-analyses. Moreover, it would be interesting to investigate whether children with DLD have a domain-general deficit in statistical learning or solely a verbal statistical learning deficit, for example by a comparison between visual and verbal distributional learning. The relationship between statistical learning and lexical-semantic knowledge should be examined further. It could be fruitful to focus on a subgroup of children with DLD who show apparent difficulties with lexical-semantic skills. Finally, measuring statistical learning on-line could be beneficial for both group comparisons as well as studying individual differences.

Chapter 6

General discussion

The language difficulties of children with developmental language disorder (DLD) include problems with understanding and expressing meaning, which could (partly) be explained by less sensitivity to statistical regularities. This dissertation aimed to investigate whether a deficit in various types of statistical learning exists in these children and whether these statistical learning abilities correlate with their lexical-semantic knowledge. Various statistical learning tasks were developed, each of which we hypothesized to measure an ability that contributes to a stage of word learning. These types of statistical learning were tested in children with and without DLD, and lexical-semantic knowledge was assessed using four standardized tasks. Furthermore, another aim of this dissertation was to find on-line measures of statistical learning that are suitable to test statistical learning ability in school-aged children with and without DLD. This chapter offers a summary and interpretation of the results found in the different studies that are reported in this dissertation.

6.1 Summary of the findings

The main aims of Chapter 2 were to compare children with and without DLD on a word segmentation task, which tests the ability of learning word boundaries in an uninterrupted stream of syllables, and to investigate whether this skill is related to children's lexical-semantic knowledge. Moreover, we aimed to develop a reliable on-line measure of word segmentation ability, as off-line measures are often difficult for children and performance is likely influenced by several cognitive factors. We recorded a number of syllables that were used to construct a unique stream for every participant. In the stream, four two-syllabic words were repeated in a random order, without any pauses or prosodic cues for word boundaries. Click sounds were inserted in the stream at random positions, either within words or between two words (Franco et al., 2015; Gómez et al., 2011). Participants were prompted to push a button as soon as they heard a click sound. We hypothesized that when participants had become sensitive to the word boundaries in the stream, they would respond faster to click sounds between words than to click sounds within words, as a click sound within a word would be more surprising. In the process of developing this task, we unexpectedly were not able to detect learning in different groups of pilot participants. TD adults and children did not show evidence of on-line or off-line learning in different versions of the task. Unfortunately, for this reason we were not able to compare word segmentation ability in children with and without DLD. The chapter reports the performance of different groups of participants in the different versions of the task that were developed and discusses possible causes for these null results.

Linking words to referents is another word learning process that has been shown to be supported by statistical learning mechanisms (see §3.2). The study reported in Chapter 3 tested implicit cross-situational word learning ability in children with and without DLD, with the use of off-line and on-line measures of learning, and investigated whether this ability is related to lexical-semantic knowledge in children. During the task, children were exposed to ambiguous trials containing novel words and novel referents, without explicit instructions. As it was not indicated which word corresponded to which referent, the correct word–referent mappings could be learned only across trials. Results from the off-line test phase that was administered after the exposure phase show that children with DLD are less proficient in learning the links between words and visual referents compared to TD children. However, the fact that both groups performed significantly above chance level indicates that children with DLD are able to learn on this task. Our eye-tracking data show some evidence for on-line learning, as participants on average looked significantly more often towards the target picture than to the distractor picture during the familiarization phase. However, we did not find evidence that this preference for the target picture increased as the experiment progressed. The data does not conclusively point to a difference between TD children and children with DLD, thus we cannot conclude whether children with DLD have a different implicit cross-situational word learning trajectory compared to TD children. Finally, our regression analyses did not reveal evidence for or against a link between implicit cross-situational word learning and lexical-semantic knowledge in children with DLD.

Research has shown that distributional learning underlies the categorization of verbal and visual input (see §4.1 and §5.2). As creating novel categories on the basis of visual input could be important for the acquisition of a rich lexicon, we wanted to test whether children with DLD are compromised in this ability. The main aim of the study reported in Chapter 4 was to develop a task that was suitable to test visual distributional learning in school-aged children with and without DLD. Our task was an adaption of the task reported by Junge et al. (2018), who tested visual distributional learning in infants. The test phase was adapted so that it was suitable for older children. Moreover, we aimed to work around a possible confound in the original study (and other distributional learning studies) by changing the design of the exposure phase, based on the study of Chládková et al. (2022). During exposure, children saw tokens from an 11-step continuum ranging from one novel animate object to another. There were two conditions (between-subjects) that differed in the frequency distribution of the tokens, which was hypothesized to influence categorization of the stimuli. A categorization task was administered after exposure. The distributional properties of the familiarization condition significantly influenced categorization during the test phase, indicating that school-aged TD children are able to form novel visual categories on the basis of distributional information. This task thus seems suitable for testing visual distributional learning in different types of participant groups.

In Chapter 5, we tested whether children with DLD show a deficit in visual distributional learning, using the task that is described in the previous chapter, and whether this ability is related to their lexical-semantic proficiency. We found that children seem to learn novel categories overall; we did not find evidence for a difference in visual distributional learning ability between children with and without DLD, nor a relationship between this ability and children with DLD's lexical-semantic skills.

6.2 Discussion

6.2.1 Aim 1: Statistical learning deficit in DLD

The first aim of this dissertation was to investigate the hypothesized statistical learning deficit in children with DLD using various statistical learning tasks. For this end, we developed three statistical learning tasks that are hypothesized to underlie different stages of word learning. Firstly, the word segmentation task (Chapter 2) mimics the process of finding words in the speech stream on the basis of statistical regularities (transitional probabilities between syllables). Second, the cross-situational word learning task (Chapter 3) was designed to simulate the learning of links between words and referents on the basis of statistical regularities (co-occurrence information). Finally, the visual distributional learning task (Chapter 4 and Chapter 5) measures categorization of new visual stimuli on the basis of statistical regularities (distributional information). In the study reported in Chapter 2, we were unable to test the statistical learning deficit hypothesis, as we did not detect learning during our word segmentation task in different groups of TD adults and TD children. In Chapter 3, we report a study that provides evidence for a deficit in implicit cross-situational word learning in children with DLD. In the study reported in Chapter 5, we do not find evidence for or against a deficit in visual distributional learning in children with DLD.

This dissertation thus provides some evidence that children with DLD are less sensitive to statistical regularities than TD children: on an off-line measure of implicit cross-situational word learning, children with DLD perform more poorly than TD children. At the same time, we did find significantly better than chance performance on this task in our group of children with DLD. Thus, children with DLD seem to be able to learn word–referent pairs implicitly, but they might need more input to get to the same level as TD children. Together with previous work into cross-situational word learning in DLD (Ahufinger et al., 2021; McGregor et al., 2022), our results suggest that the statistical learning deficit in children with DLD that is documented for sequential statistical learning extends to learning co-occurrences between words and objects. New in our study is the use of an implicit task. In both previous studies (Ahufinger et al., 2021; McGregor et al., 2022), children were explicitly instructed to learn new names of new objects, and in the case of Ahufinger et al. (2021) wen through a practice phase before being subjected to the exposure phase. In our study, the children were instructed only to pay attention to the pictures and the words, and did a cover task during the exposure phase. This is more in line with usual investigations of statistical learning, in which it is tested whether passive exposure to statistical regularities results in implicit knowledge of those regularities. We hypothesize that if children with DLD are less able to utilize co-occurrences between words and visual referents in their environment, this could greatly hamper the building of a rich lexicon. However, we cannot underpin this hypothesis with evidence for a relationship between cross-situational word learning ability and existing lexical-semantic knowledge in children with DLD (see §6.2.2).

While we found evidence for poorer implicit cross-situational word learning on the off-line test phase in Chapter 3, the remaining research questions belonging to the first research aim rendered null results. Our on-line measure (eye-tracking data) of cross-situational word learning did not reveal evidence of divergent patterns in looking behaviour in our children with DLD compared to our TD children. We thus have no evidence that children with DLD have a different cross-situational word learning trajectory. Furthermore, we report no evidence for a deficit in visual distributional learning in children with DLD in **Chapter 5**. Overall, our participants showed evidence for learning on the visual distributional learning task. Nevertheless, we cannot be certain whether this means that children with DLD are able to exploit visual distributional learning mechanisms; an exploratory analysis on the data from the DLD group separately did not reveal performance that was significantly better than chance. We thus cannot conclude whether visual distributional learning is affected or intact in children with DLD.

The fact that we did not directly statistically compare performance on the cross-situational word learning task (Chapter 3) and the visual distributional learning task (Chapter 5), means we cannot conclude that children with DLD performed better or worse on one task as opposed to the other. Thus, we can only speculate about why we found evidence for a deficit in cross-situational word learning while the results for visual distributional learning are inconclusive. The tasks differ in domain: the cross-situational word learning task is a visual-auditory verbal task, while the visual distributional learning task is a visual non-verbal task. Theoretically, it is possible that a statistical learning deficit in DLD is limited to verbal statistical learning, while visual statistical learning is (relatively) intact. The results of the different studies in the dissertation of Lammertink (2020), although also inconclusive, seem to point into this direction as well. In her dissertation, she reports a (small) deficit in verbal statistical learning (non-adjacent dependency learning; Lammertink et al., 2019) in children with DLD, but intact visuo-motoric non-verbal statistical learning (Lammertink, Boersma, Wijnen, et al., 2020) and visual non-verbal statistical learning (Lammertink, Boersma, Rispens, et al., 2020). Null results considering a deficit in visual statistical learning are reported by Noonan (2018). A specific verbal statistical learning deficit cannot be concluded from these studies either; it is not possible to make a direct comparison between performance on the different statistical learning tasks, as they all target different types of underlying statistical structures. There has been some evidence in the literature for a deficit in visual statistical learning in children with DLD (Collisson et al., 2015; Gillis et al., 2022; Lukács et al., 2021); however, it remains to be seen whether a meta-analysis will lead to a significant difference between individuals with and without DLD when all studies are pooled. The question thus remains whether the statistical learning deficit in DLD, assuming that it exists, is domain-general or domain-specific (Arciuli & Conway, 2018). In fact, there has been debate about the domain-generality of statistical learning mechanisms in general (Arciuli, 2017; Frost et al., 2015; Siegelman, 2020). Besides differing in domain, the statistical learning tasks that are reported in this dissertation also vary in the type of statistical regularities that are presented to the children (co-occurrences, frequency distribution). To get a more specific picture of the statistical learning deficit in children with DLD, systematic comparisons between performance on statistical learning tasks in different domains, modalities and with different underlying statistical regularities should be made. Meta-analyses of statistical learning ability of people with DLD have shown that they likely have a deficit in different types of statistical learning (Lammertink et al., 2017; Lum et al., 2014; Lum & Conti-Ramsden, 2013; Obeid et al., 2016), but have not focused on visual statistical learning. Such a study could provide insight into the probability and size of a potential visual statistical learning deficit in individuals with DLD.

Our finding of a deficit in implicit cross-situational word learning in children with DLD is most likely caused by less sensitivity to statistical regularities, but it is possible that performance on this task is (also) influenced by cognitive abilities besides statistical learning, such as working memory and attention. Although we aimed to reduce the difficulty in differentiating between newly learned words by conducting novel words that varied in phonological structure (CVCV, CVCVC, CVC), and included a higher number of exposures (seven) compared to Ahufinger et al. (2021; four exposures), difficulty with phonological processing might still have hampered learning in the children with DLD. Roembke et al. (2020) showed that cross-situational word learning ability is affected by reduced working memory recourses. Unfortunately, we were not able to control for phonological processing abilities in our statistical analyses, as the background measures were assessed only in our group of children with DLD, and not the TD children[16].

Difficulties with attention could also have influenced performance negatively in the children with DLD. The eye-tracking data gave us hints for that idea. We found a clear imbalance in the number of data points: the TD children provided a lot more data than the children with DLD (see Appendix 1 and Appendix 2 for graphs), partly because in the DLD group there were a lot more “irrelevant” looks (meaning a child was looking at

[16] Due to the outbreak of the covid-19 pandemic in 2020, the primary schools in the Netherlands were closed for a considerable part of the school year. Therefore we were unable to test a new group of TD children on the statistical learning tasks and background measures.

the screen, but not at either of the two images). We cannot be certain whether this means that the children with DLD had less access to the stimuli, directly hampering word learning, or that differences in attentional abilities influenced the statistical learning process during the task. Previous research has shown that attentional abilities affect statistical learning processes (Toro et al., 2005; West et al., 2021; Yu et al., 2012), and that attention is affected in DLD (Ebert & Kohnert, 2011; Smolak et al., 2020). McGregor et al. (2022) report that sustained attention ability seems to play a role in cross-situational word learning in children with DLD. Unfortunately, we have not measured attention in our participants. Still, if it were the case that attention and/or working memory abilities influenced statistical learning performance in our cross-situational word learning task, this would not necessarily be an argument against the statistical learning deficit hypothesis. As mentioned in the previous paragraph, it has been proposed that statistical learning is not one unified ability, but rather is composed of different cognitive abilities (the "multicomponent view of statistical learning") such as processing speed, memory and attention (Arciuli, 2017; Frost et al., 2015, 2019; Siegelman, Bogaerts, Christiansen, et al., 2017), but see Bogaerts et al. (2022) for arguments in favour of the existence of a general statistical learning ability). Future research should determine whether statistical learning indeed consists of different components, and if that is the case, to what extent they are affected in DLD. In a more practical sense, it would be useful to find ways to keep the children interested in the task, for example with the use of moving stimuli.

Our statistical models yielded wide confidence intervals, because of the heterogeneity of the DLD population and the relatively low number of participants. This makes it impossible to conclude whether an underlying effect could be small or large. Testing larger groups of participants will narrow the confidence interval, thereby raising the chance that it does not contain zero (if a true effect exists in the population) or that we can conclude that the true effect is small or non-existent (if a true effect does not exist). Our initial goal was to include 50 participants per group, but unfortunately, we were not able to achieve this due to the outbreak of covid-19. This could have especially influenced the width of the confidence interval of the Group difference in the study described in Chapter 5, because of the between-participant design. In this study, a learning effect could only be detected by the finding of a significant difference in performance in the two familiarization conditions. A deficit in distributional learning in children with DLD, could thus only be tested by looking at the interaction between Condition and Group. This resulted in fairly small subgroups of participants. The between-participant design could possibly be avoided by including a pre-test and a post-test of categorization (Chládková & Šimáčková, 2021).

While the off-line measure of our cross-situational word learning task revealed a significant difference between our children with and without DLD, the on-line measure did not. We expected to find that children with DLD look significantly less often towards the target picture than TD children. Moreover, we expected to find an on-line learning effect, entailing that the proportion of looks towards the target picture (as opposed to the distractor picture) would increase in the course of the experiment. This on-line learning effect was hypothesized to be weaker in children with DLD than TD children. However, these effects were not statistically significant. Here we can speculate a bit more about some exploratory findings that can be observed in the data. For analysis, the eye-tracking data was split up into two time windows: Word1 and Word2, corresponding to the first word that was played in a learning trial and the second word. When only Word2 was taken into account, there are some differences in the expected direction: TD children look more towards the target picture overall, compared to children with DLD (main effect of Group, 95% CI in odds ratio: 0.94 .. 2.29, p = 0.086), and the on-line learning effect is stronger in the TD children compared to children with DLD (interaction between Trial and Group, 95% CI in odds ratio: 0.94 .. 1.73, p = 0.110). This suggests that the underlying effect might be moderate to large, but could also be non-existent. These exploratory findings are interesting to investigate in future research.

6.2.2 Aim 2: relationship SL and lexical-semantic proficiency

We theorized that children with DLD have a deficit in statistical learning, which contributes to their lexical-semantic difficulties. Therefore, the second aim of this dissertation was to investigate whether individual differences in different types of statistical learning are related to individual differences in various lexical-semantic skills in children with and without DLD. Unfortunately, we were not able to test this association in TD children[16]. Lexical-semantic knowledge was measured using several tasks, tapping passive vocabulary size, active vocabulary size, word category knowledge and lexical-semantic organization. This is a more extensive test battery of lexical(-semantic) knowledge than reported in previous studies. To investigate this relationship, we ran multiple linear mixed effect models in which we tested the influence of cross-situational word learning ability (both the off-line and on-line measure) and visual distributional learning ability on the lexical-semantic abilities in children with DLD, taking into account the variation in age, social-economic status, phonological processing, verbal working memory and non-verbal intelligence.

In none of the studies that are reported in this dissertation, we find evidence for or against a link between statistical learning and language. Unfortunately, this null result is difficult to interpret; it does not mean that we can conclude that the association is absent. Significant correlations between statistical learning ability and language proficiency have been reported for TD individuals (Conway et al., 2010; Ellis et al., 2014; Evans et al., 2022; Gerbrand et al., 2022; Hamrick et al., 2018a; Isbilen et al., 2022; Kaufman et al., 2010; Kautto & Mainela-Arnold, 2022; Kemény & Lukács, 2021; Kidd, 2012; Kidd & Arciuli, 2016; McGregor et al., 2022; Misyak et al., 2010; Newman et al., 2006; Shafto et al., 2012; Spencer et al., 2015; Vlach & DeBrock, 2017) and children with DLD (Ahufinger et al., 2022; Evans et al., 2009; Hedenius et al., 2011; Mainela-Arnold & Evans, 2014; Misyak et al., 2010; Sack et al., 2021; Tomblin et al., 2007). However, null results also have been reported in multiple articles (Aguilar & Plante, 2014; Lammertink, Boersma, Rispens, et al., 2020; Lammertink, Boersma, Wijnen, et al., 2020; Lammertink et al., 2019; Noonan, 2018). Moreover, when significant correlations are reported, the values often indicate a weak relationship, especially in children (Arnon, 2019). We can only speculate about possible causes of our results. The wide confidence intervals (see §3.5.3 and §5.5.3) show that the true underlying effects might be very small, very large or anything in between. As we could not include our group of TD children in the analysis, we had data from only 25 children. Future studies, besides aiming to test larger groups of participants, could set up longitudinal research designs to test the relationship between statistical learning ability and lexical-semantic knowledge in children with and without DLD.

It is possible that performance on statistical learning tasks is not a reliable indicator of individual differences in statistical learning ability. Statistical learning tasks have been designed to measure group-level performance, and the psychometric probabilities of the task might not be adequate to measure individual differences in statistical learning ability reliably (Arnon, 2019; Siegelman, Bogaerts, & Frost, 2017; Siegelman, Bogaerts, Christiansen, et al., 2017). Arnon (2019) tested adults and children on three different statistical learning tasks that are commonly used for measuring individual differences (two auditory tasks and one visual task) at two moments in time. The reliability of the tasks for adults was moderate (albeit lower than psychometric norms), but for children both test-retest reliability and internal consistency (indicating how well different test items measure the same underlying construct) was low, and there were no significant correlations between performances on the different tasks. Possible causes of low reliability could be the relatively small number of test items in which stimuli are often repeated, and the lack of variation in difficulty of the test items. Moreover, often many participants perform around chance level, indicating that they are guessing, causing their scores to be noisy rather than informative and individual variation to be low. Finally, test questions are usually quite explicit while learning on the statistical learning task is likely (more) implicit. Importantly, all these factors could affect children more strongly than adults. If performance on a statistical learning task does not capture individual statistical learning abilities reliably, it is not meaningful to use these values to predict language proficiency; doing so could lead to both overestimation and underestimation of the true effect (Siegelman, Bogaerts, & Frost, 2017). This is an important problem in the field of statistical learning and more research should be executed to find reliable measures of statistical learning, especially in children. Possible solutions are the use of on-line measures, increasing the number of test questions and adding different levels of difficulty in them (Siegelman, Bogaerts, & Frost, 2017), or making the test more implicit. We aimed to test statistical learning more reliably by including on-line measures of learning, but unfortunately we did not find strong evidence for these measures reflecting on-line learning. Although the split-half reliability of our visual distributional learning task (Chapter 4 and Chapter 5) is $r = .73$, which is higher than what Arnon (2019) found for her child participants, the test–retest reliability is unknown.

Although we cannot conclude this on the basis of our null results, it is also possible that lexical-semantic knowledge is not strongly supported by statistical learning mechanisms, and a deficit in statistical learning is not an explanation for the lexical-semantic deficits in DLD. Statistical learning abilities may be more strongly related to language areas that are more sequential in nature, while declarative memory is more important for the development of lexical-semantic knowledge (Mainela-Arnold & Evans, 2014). It has also been argued that declarative memory plays a compensatory role in children with DLD (Ullman & Pullman, 2015). For example, on the basis of Pearson correlations, Lum et al. (2012) report that procedural memory is related to grammar while declarative memory is related to vocabulary in TD children. On the other hand, the grammar of children with DLD seems to be supported by declarative memory and not procedural memory. However, a direct comparison between the *r*-values of the correlations did not yield a significant *p*-value (0.057). Similarly, Kemény and Lukács (2021) report that statistical learning ability contributes to vocabulary in TD children, while this is not the case for children with DLD, and subsequently argue that this might point to a difference in the cognitive processes underlying language acquisition in the two populations. It is important to note that Kemény and Lukács (2021) make no direct statistical comparison between TD children and children with DLD. In that case, it is not valid to conclude that groups are different on the basis of a difference in *p*-values (Nieuwenhuis et al., 2011). Importantly, McGregor et al. (2022) report that vocabulary measures contribute more to cross-situational word learning ability in TD children than in children with DLD, and suggest this could be due to a compensatory role for declarative memory in children with DLD. Potential differences in this sense between children with DLD and TD children could not be investigated with the use of our data, neither were we able to contrast statistical learning to declarative memory. In any case, this dissertation has not been able to provide evidence for or against a statistical learning deficit underlying the lexical-semantic difficulties in children with DLD. An alternative hypothesis is that differences in attention and phonological working memory abilities in children with DLD contribute to their lexical-semantic difficulties (Mainela-Arnold et al., 2010).

6.2.3 Aim 3: on-line measures of statistical learning

As discussed in the previous section, off-line measures of statistical learning are likely not always a reliable indicator of individual statistical learning abilities, especially in children. On the other hand, on-line measures are potentially more sensitive measures of statistical learning and can provide insight into the learning process itself, instead of only into the end product. Therefore, the third aim of this dissertation was to develop on-line measures of statistical learning. In Chapter 2 we report our attempt to measure word segmentation ability on-line using a click detection task. In Chapter 3, we report the use of eye-tracking in our cross-situational word learning task. Unfortunately, we did not find conclusive evidence that these measures reflect statistical learning processes, although our eye-tracking data did reflect that children look more towards the target picture than the competitor overall. We did not include an on-line measure in the visual distributional learning task reported in Chapter 4 and Chapter 5. To the best of our knowledge, this has not been reported in other distributional learning studies. Future studies should try to find a way to measure distributional learning on-line, perhaps by inserting test questions at different moments in the exposure phase.

In Chapter 2, we implemented the click detection task as a measure of on-line word segmentation ability (see also Franco et al., 2015; Gómez et al., 2011). At the time, there were no studies of on-line measures of word segmentation in children. As a first pilot study, we tested our child-friendly word segmentation task in adults. Unfortunately, we cannot conclude that the click detection task measures on-line word segmentation ability. We expected that when participants become sensitive to the word boundaries in the speech stream, clicks within words would be harder to detect and thus cause slower reaction times as opposed to clicks between words. However, we did not find evidence for an influence of the position of a click (between words vs. within words) on reaction time, nor for the effect of click position increasing as the experiment progressed (i.e. interaction between Click position and Block). We did find a significant three-way interaction between Click position, Block and Version, indicating that the effect of click position across blocks is different for the two versions of the task. Upon visual inspection of the data (meaning we cannot say anything about the statistical significance of the effects), the participants that did version A of the experiment behaved more like we expected than the participants that did version B of the experiment: they were faster at detecting clicks between words and this effect increased across blocks. However, participants that did version B of the experiment seemed to show the opposite effect in later blocks. When we inspected individual data, we found that there was a large amount of individual variation. Thus, while the response time to clicks might be sensitive to word boundary knowledge in some participants, it does not seem to be a reliable measure of on-line statistical learning at the group level. As we discuss in Chapter 2, it might be the case that the addition of the click detection task hampered statistical learning. Moreover, the "auditory streaming effect" (Micheyl et al., 2010; van Noorden, 1975), might have caused the participants to hear the syllables and the click sounds as two separate streams, and therefore measuring reaction times to click sounds is not a reliable measure sensitivity to the word boundaries.

Promising results have been reported for measuring word segmentation on-line with the use of a syllable detection task (Lukics & Lukács, 2021). In that study, adult participants were prompted to press a key as fast as possible when they heard a specific syllable, which was the final syllable of one of the words that was repeated throughout the exposure phase. After three blocks, a random block was inserted, during which the syllable was not predictable, followed by a recovery block during which the target syllable was predictable again. Participants were significantly slower at detecting the target in the random block and then significantly improved again in the recovery block, indicating that arising knowledge of the transitional probabilities between syllables influences ease of response to a target syllable when it is predictable. The authors furthermore report that the on-line scores significantly correlated with performance on the off-line forced choice task, and that internal consistency was higher for the on-line measure. This task has also been used to compare statistical learning ability between children with and without DLD (Lukács et al., 2021), and their results indicate that the syllable detection task is a sensitive measure of on-line word segmentation ability. Future research should be done to investigate whether individual differences in on-line word segmentation ability correlate with measures of lexical-semantic knowledge in children with and without DLD.

As an on-line measure of implicit cross-situational word learning, we collected eye gaze data during the exposure phase of the task reported in Chapter 3. Apart from differences in looking behaviour between TD children and children with DLD, we expected to see evidence of on-line learning. Specifically, we expected that if children were learning the links between words and referents during the course of the experiment, they would look more towards the correct picture at the end of the experiment compared to the start of the experiment. That is, we expected a main effect of Trial on the proportion of looks towards the target. As a sanity check, we computed this effect, as well as the significance of the Intercept of our model. An Intercept significantly higher than 0 would indicate that overall children look more towards the target picture than towards the distractor picture. However, our expectations were only partly met. The effect of

Trial was not significant, and the Intercept was only significant in the model of Word2. This final finding does indicate that children's looking behaviour is influenced by developing knowledge of word–referent pairs.

Measuring eye gaze as an index for processing information is a quite well-established method. Specifically for cross-situational word learning experiments, it has been reported that looking behaviour seems to reflect on-line learning in adults (Yu et al., 2012) and infants (Yu & Smith, 2011). Ahufinger et al. (2021) find no evidence for a difference in on-line cross-situational word learning between TD children and children with DLD. In fact, they do not find evidence that eye-gaze reflects learning of word–referent pairs overall, as their participants showed no significant preference for the target picture as opposed to the distractor picture. As discussed above in §6.2.1, it could be the case that differences in attentional abilities influence looking behaviour strongly in children with DLD. More work is needed to find reliable measures of on-line cross-situational word learning in children with and without DLD. It is also possible to insert test questions in the exposure phase of the cross-situational word learning task, enabling the tracking of word–referent knowledge. However, that would make such a task less implicit. Using eye-tracking has the advantage of not requiring participants to do an additional task but still providing a fine-grained behavioural measure. Our results, although inconclusive, do suggest that during the process of learning word–referent knowledge looking behaviour changes, so we believe it is worthwhile to experiment with this method in children with and without DLD.

6.3 Recommendations for future research and clinical implications

Moving forward in the investigation of statistical learning ability in individuals with DLD, it would be wise to set up experiments that are more strongly theoretically justified. Bogaerts et al. (2021) argue that research into this area has often been vague about the theoretical constructs of statistical learning that they assume (for example, terms as statistical learning, implicit learning, sequential learning and procedural learning are used in research, but often not clearly defined or distinguished from each other), how these constructs relate to different types of statistical learning tasks and how those tasks relate to the language difficulties in DLD. For example, statistical learning in individuals with DLD has been investigated using a wide arrange of tasks, and it is often not specified how exactly the statistical regularities in the particular task correspond to language problems in DLD. Often, such studies only include one confirmatory hypothesis. Bogaerts et al. (2021) state that it is important to set up experiments that contain specific predictions about which abilities are impaired in children in DLD, and which are intact ("a parallel test of exclusion", p. 12). They argue that to show that a particular population has a deficit in a certain theoretical construct, in this case statistical learning, findings of poor performance of that construct should be contrasted with findings of intact performance on a task that does not require that specific construct. For example, finding poor visual statistical learning in DLD contrasted with normal sensory processing of similar visual input, would be a stronger argument for the statistical learning deficit hypothesis than merely finding impaired performance on a visual statistical learning task. Although for this dissertation we have tried to construct statistical learning tasks that theoretically could underlie different processes in the targeted ability (lexical-semantic knowledge), this could be done with more theoretical precision in future research, and adding a test of exclusion would be insightful.

Concerning the investigation of the relationship between statistical learning and language proficiency, future studies could aim to conduct statistical learning tasks that are more ecologically valid. While statistical learning tasks are strongly simplified compared to real-life statistics, learning on these types of tasks is still expected to predict performance on broad language tasks that require sensitivity to a range of different types of regularities (Isbilen et al., 2022). Bogaerts et al. (2021) argue that statistical learning tasks should reflect the statistical regularities that exist in different types of input more closely, for example through analysis of big data or corpus data. With the goal of clarifying the link between statistical learning and language, Isbilen et al. (2022) made a comparison between artificial language learning and natural language learning of similar structures. They found that performance on a word segmentation task in which it is tested whether participants become sensitive to trisyllabic words predicts sensitivity to similar statistical patterns in natural language. Such ecologically valid statistical learning tasks might be more suitable for establishing associations between statistical learning and language in children with and without DLD.

It is important to connect experimental findings to clinical implications for children with DLD. Nevertheless, it is difficult to state clear recommendations on the basis of the current dissertation, as we did not conduct intervention studies. Interventions based on statistical learning as a mechanism underlying language seem to be successful for children with DLD (Plante & Gómez, 2018). As our study reported in Chapter 3 indicates that implicit cross-situational word learning is affected in children with DLD, it might be useful to focus on improvement of that skill in language therapies. The principles of cross-situational word learning have been applied in small-scale intervention studies in children with small expressive vocabularies (Alt et al., 2014; Ng et al., 2020). In those studies, a list of 5 to 10 unknown words was composed separately for each child, together with a set of control words. In 14 to 20 sessions, target words, but not control words, were presented to the children at least 64 times in different contexts. The number of presentations of the target words was automatically kept up with the use of a data tracker. All children started to use the target words significantly more often than the control words, but also showed a significant increase in vocabulary size overall, indicating that high contextual diversity and high dosage enhances word learning in children with DLD, and this effect generalizes to general word learning ability. In accordance with these findings, our results indicate that children with DLD are able to learn word–referent pairs in an implicit task, but need more input than TD children. Together, these results imply that providing more and diverse input in language therapy is fruitful for (implicit) word learning in children with DLD. This should be tested in larger-scale intervention studies.

6.4 Conclusion

The aim of this dissertation was to investigate the relationship between different types of statistical learning and lexical-semantic knowledge in children with and without DLD, as well as finding suitable on-line measures of statistical learning. Implicitly keeping track of the co-occurrences between words and what they refer to could play an important role in building an elaborate mental lexicon. Our results indicate that children with DLD have more difficulty than TD children with learning to couple words to their referents in situations with multiple words and multiple potential referents, which could be due to a reduced statistical learning ability. This is an extension of previously reported results: besides sequential statistical learning, statistical learning of co-occurrence information also seems to be affected in DLD, at least when it concerns verbal-visual input. We did not find evidence for or against a deficit in visual distributional learning in DLD nor a relationship between statistical learning and lexical-semantic knowledge. More research into the underlying causes of these lexical-semantic difficulties in DLD is necessary. Statistical learning tasks that are more ecologically valid and are complemented with a test of exclusion, as well as increasing reliability with the use of on-line measures and/or more elaborate test phases with test items that vary in difficulty, may be the way to move forward in investigating statistical learning and its relation to language in children with and without DLD.

References

Aguilar, J. M., & Plante, E. (2014). Learning of grammar-like visual sequences by adults with and without language-learning disabilities. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1394–1404. https://doi.org/10.1044/2014_JSLHR-L-13-0124

Aguilar, J. M., Plante, E., & Sandoval, M. (2017). Exemplar variability facilitates retention of word learning by children with specific language impairment. *Language, Speech, and Hearing Services in Schools*, *49*(1), 1–13. https://doi.org/10.23641/asha

Ahufinger, N., Ferinu, L., Sanz-Torrent, M., Andreu, L., & Evans, J. L. (2022). Statistical word learning in Catalan-Spanish and English-speaking children with and without developmental language disorder. *International Journal of Language & Communication Disorders*, *57*(1), 42–62. https://doi.org/10.1111/1460-6984.12673

Ahufinger, N., Guerra, E., Ferinu, L., Andreu, L., & Sanz-Torrent, M. (2021). Cross-situational statistical learning in children with developmental language disorder. *Language, Cognition and Neuroscience*, *36*(9), 1180–1200. https://doi.org/10.1080/23273798.2021.1922723

Alt, M., Meyers, C., Oglivie, T., Nicholas, K., & Arizmendi, G. (2014). Cross-situational statistically based word learning intervention for late-talking toddlers. *Journal of Communication Disorders*, *52*, 207–220. https://doi.org/10.1016/J.JCOMDIS.2014.07.002

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *49*(5), 941–954. https://doi.org/10.1044/1092-4388(2006/068)

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research*, *47*(2), 407–420. https://doi.org/10.1044/1092-4388(2004/033)

Altvater-Mackensen, N., Jessen, S., & Grossmann, T. (2017). Brain responses reveal that infants' face discrimination is guided by statistical learning from distributional information. *Developmental Science*, *20*(2). https://doi.org/10.1111/desc.12393

Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160058. https://doi.org/10.1098/RSTB.2016.0058

Arciuli, J., & Conway, C. M. (2018). The promise—and challenge—of statistical learning for elucidating atypical language development. *Current Directions in Psychological Science*, *27*(6), 492–500. https://doi.org/10.1177/0963721418779977

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464–473. https://doi.org/10.1111/j.1467-7687.2009.00937.x

Arnon, I. (2019). Statistical learning, implicit learning, and first language acquisition: a critical evaluation of two developmental predictions. *Topics in Cognitive Science*, *11*(3), 504–519. https://doi.org/10.1111/tops.12428

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*, *52*(1), 68–81. https://doi.org/10.3758/s13428-019-01205-5

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*(5), 875–890. https://doi.org/10.1111/J.1467-8624.1991.TB01577.X

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/JSS.V067.I01

Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex*, *115*, 56–71. https://doi.org/10.1016/j.cortex.2019.01.013

Beckage, N. M., Smith, L. B., & Hills, T. (2010). Semantic network connectivity is related to vocabulary growth rate in children. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2769–2774.

Bishop, D. V. (2003). Specific language impairment: diagnostic dilemmas. In *Classification of Developmental Language Disorders* (pp. 321–338). Psychology Press. https://doi.org/10.4324/9781410609021-21

Bishop, D. V. (2006). What causes specific language impairment in children? *Current Directions in Psychological Science*, *15*(5), 217–221. https://doi.org/10.1111/j.1467-8721.2006.00439.x

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Adams, C., Archibald, L., Baird, G., Bauer, A., Bellair, J., Boyle, C., Brownlie, E., Carter, G., Clark, B., Clegg, J., Cohen, N., Conti-Ramsden, G., Dockrell, J., Dunn, J., Ebbels, S., … Whitehouse, 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. https://doi.org/10.1111/JCPP.12721

Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: a meta-analysis. *39th Annual Meeting of the Cognitive Science Society*, 124–129.

Blythe, R. A., Smith, K., & Smith, A. D. M. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science*, *34*(4), 620–642. https://doi.org/10.1111/j.1551-6709.2009.01089.x

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer* (6.1.05). http://www.praat.org/

Bogaerts, L., Siegelman, N., Christiansen, M. H., & Frost, R. (2022). Is there such a thing as a 'good statistical learner'? *Trends in Cognitive Sciences*, *26*(1), 25–37. https://doi.org/10.1016/J.TICS.2021.10.012

Bogaerts, L., Siegelman, N., & Frost, R. (2021). Statistical learning and language impairments: toward more precise theoretical accounts. *Perspectives on Psychological Science*, *16*(2), 319–337. https://doi.org/10.1177/1745691620953082

Borovsky, A., Burns, E., Elman, J. L., & Evans, J. L. (2013). Lexical activation during sentence comprehension in adolescents with history of specific language impairment. *Journal of Communication Disorders*, *46*(5–6), 413–427. https://doi.org/10.1016/j.jcomdis.2013.09.001

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. https://doi.org/10.1111/1460-6984.12518

Brackenbury, T., & Pye, C. (2005). Semantic deficits in children with language impairments: issues for clinical assessment. *Language, Speech, and Hearing Services in Schools*, *36*(5), 5–16.

Broedelet, I., Boersma, P., & Rispens, J. (2021). Measuring (online) word segmentation in adults and children. *Dutch Journal of Applied Linguistics*, 10. https://doi.org/10.51751/dujal9607

Broedelet, I., Boersma, P., & Rispens, J. (2022). School-aged children learn novel categories on the basis of distributional information. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.799241

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). MIT Press.

Charest, M., & Skoczylas, M. J. (2019). Lexical diversity versus lexical error in the language transcripts of children with developmental language disorder: Different conclusions about lexical ability. *American Journal of Speech-Language Pathology*, *28*(3), 1275–1282. https://doi.org/10.1044/2019_AJSLP-18-0143

Chládková, K., Boersma, P., & Escudero, P. (2022). Unattended distributional training can shift phoneme boundaries. *Bilingualism: Language and Cognition*, 1–14. https://doi.org/10.1017/S1366728922000086

Chládková, K., & Šimáčková, Š. (2021). Distributional learning of speech sounds: an exploratory study into the effects of prior language experience. *Language Learning*, *71*(1), 131–161. https://doi.org/10.1111/LANG.12432

Chung, H., & Yim, D. (2020). Quick incidental learning of words by children with and without specific language impairment: an eye-tracking study. *Communication Sciences and Disorders*, *25*(3), 499–516. https://doi.org/10.12963/CSD.20715

Cohen, L., & Mehler, J. (1996). Click monitoring revisited: An on-line study of sentence comprehension. *Memory & Cognition 1996 24:1*, *24*(1), 94–102. https://doi.org/10.3758/BF03197275

Cole, R. A. (1980). Perception and production of fluent speech. In *Perception and Production of Fluent Speech* (1st ed.). Taylor and Francis. https://doi.org/10.4324/9781315638935

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin and Review*, *21*(4), 843–860. https://doi.org/10.3758/s13423-013-0564-3

Collisson, B. A., Grela, B., Spaulding, T., Rueckl, J. G., & Magnuson, J. S. (2015). Individual differences in the shape bias in preschool children with specific language impairment and typical language development: Theoretical and clinical implications. *Developmental Science*, *18*(3), 373–388. https://doi.org/10.1111/desc.12219

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356–371. https://doi.org/10.1016/j.cognition.2009.10.009

Desmottes, L., Maillart, C., & Meulemans, T. (2017). Mirror-drawing skill in children with specific language impairment: Improving generalization by incorporating variability into the practice session. *Child Neuropsychology*, *23*(4), 463–482. https://doi.org/10.1080/09297049.2016.1170797

Dink, J., & Fergusson, B. (2015). *eyetrackingR: an R library for eye-tracking data analysis*. http://www.eyetrackingr.com

Dockrell, J. E., Messer, D., & George, R. (2001). Patterns of naming objects and actions in children with word finding difficulties. *Language and Cognitive Processes*, *16*(2–3), 261–286. https://doi.org/10.1080/01690960042000030

Dockrell, J. E., Messer, D., George, R., & Ralli, A. (2003). Beyond naming patterns in children with WFDs-definitions for nouns and verbs.

Journal of Neurolinguistics, *16*(2–3), 191–211. https://doi.org/10.1016/S0911-6044(02)00012-X

Dosi, I., Gavriilidou, Z., & Dourou, C. (2021). Definitional skills of learners with and without developmental language disorder. *International Journal of Learning, Teaching and Educational Research*, *20*(10), 193–216. https://doi.org/10.26803/IJLTER.20.10.11

Drljan, B., & Vuković, M. (2019). Comparison of lexical-semantic processing in children with developmental language disorder and typically developing peers. *GOVOR*, *36*(2), 119–138. https://doi.org/10.22210/govor.2019.36.07

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 and Communication Disorders*, *53*(4), 799–810. https://doi.org/10.1111/1460-6984.12385

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. https://doi.org/10.1044/1092-4388(2011/10-0231)

Elazar, A., Alhama, R. G., Bogaerts, L., Siegelman, N., Baus, C., & Frost, R. (2022). When the 'tabula' is anything but 'rasa:' what determines performance in the auditory statistical learning task? *Cognitive Science*, *46*(2). https://doi.org/10.1111/COGS.13102

Ellis, E. M., Gonzalez, M. R., & Deák, G. O. (2014). Visual prediction in infancy: what is the association with later vocabulary? *Language Learning and Development*, *10*(1), 36–50. https://doi.org/10.1080/15475441.2013.799988

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, *61*(2), 177–199. https://doi.org/10.1016/j.cogpsych.2010.05.001

Erickson, L. C., Kaschak, M. P., Thiessen, E. D., & Berry, C. A. S. (2016). Individual differences in statistical learning: conceptual and measurement issues. *Collabra*, *2*(1). https://doi.org/10.1525/collabra.41

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, *130*(4), EL206–EL212. https://doi.org/10.1121/1.3629144

Evans, J. L., Maguire, M. J., & Sizemore, M. L. (2022). Neural patterns elicited by lexical processing in adolescents with specific language impairment: support for the procedural deficit hypothesis? *Journal of Neurodevelopmental Disorders*, *14*(1), 1–22. https://doi.org/10.1186/S11689-022-09419-z

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321–335. https://doi.org/10.1044/1092-4388(2009/07-0189)

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499. https://doi.org/10.1016/j.cognition.2008.04.002

Finn, A. S., Kharitonova, M., Holtby, N., & Sheridan, M. A. (2018). Prefrontal and hippocampal structure predict statistical learning ability in early childhood. *Journal of Cognitive Neuroscience*, *31*(1), 126–137. https://doi.org/10.1162/jocn_a_01342

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, *35*(2), 367–380. https://doi.org/10.1111/j.1551-6709.2010.01156.x

Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, *4*(5), 414–420. https://doi.org/10.1016/S0022-5371(65)80081-0

Foss, D. J., & Lynch, R. H. (1969). Decision processes during sentence comprehension: Effects of surface structure on decision times. *Perception & Psychophysics 1969 5:3*, *5*(3), 145–148. https://doi.org/10.3758/BF03209545

Franco, A., Gaillard, V., Cleeremans, A., & Destrebecqz, A. (2015). Assessing segmentation processes by click detection: online measure of statistical learning, or simple interference? *Behavior Research Methods*, *47*(4), 1393–1403. https://doi.org/10.3758/s13428-014-0548-x

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/BUL0000210

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Furrer, S. D., & Younger, B. A. (2005). Beyond the distributional input? A developmental investigation of asymmetry in infants' categorization of cats and dogs. *Developmental Science*, *8*(6), 544–550. https://doi.org/10.1111/j.1467-7687.2005.00446.x

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. https://doi.org/10.1017.S0142716406060383

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, *33*(6), 966–979. https://doi.org/10.1037/0012-1649.33.6.966

Gathercole, S. E., Willis, S. W., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: a longitudinal study. *Developmental Psychology*, *28*(5), 887–898. https://doi.org/10.1037/0012-1649.28.5.887

Gerbrand, A., Gredebäck, G., Hedenius, M., Forsman, L., & Lindskog, M. (2022). Statistical learning in infancy predicts vocabulary size in toddlerhood. *Infancy*, *27*(4), 700–719. https://doi.org/10.1111/INFA.12471

Gillis, J. U., Gul, A., Fox, A., Parikh, A., & Arbel, Y. (2022). Artificial grammar learning in children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *65*(1), 292–302. https://doi.org/10.1044/2021_JSLHR-21-00050

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou & John, A., & Trueswell, C. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Gómez, D. M., Bion, R. A. H., & Mehler, J. (2011). The word segmentation process as revealed by click detection. *Language and Cognitive Processes*, *26*(2), 212–223. https://doi.org/10.1080/01690965.2010.482451

Gómez, R. L. (2002). Variability and detection of inviariant structure. *Psychological Science*, *13*(5), 431–436. https://doi.org/10.1111/1467-9280.00476

Gopnik, M. (1997). Language deficits and genetic factors. *Trends in Cognitive Sciences*, *1*(1), 5–9. https://doi.org/10.1016/S1364-6613(97)01005-X

Gordon, J., Schumm, J. S., Coffland, C., & Doucette, M. (1992). Effects of inconsiderate versus considerate text on elementary students' vocabulary learning. *Reading Psychology*, *13*(2), 157–169. https://doi.org/10.1080/027027192130204

Goujon, A. (2011). Categorical implicit learning in real-world scenes: Evidence from contextual cueing. *Quarterly Journal of Experimental Psychology*, *64*(5), 920–941. https://doi.org/10.1080/17470218.2010.526231

Graf Estes, K. (2009). From tracking statistics to learning words: Statistical learning and lexical acquisition. *Linguistics and Language Compass*, *3*(6), 1379–1389. https://doi.org/10.1111/j.1749-818X.2009.00164.x

Graf Estes, K. (2012). Infants generalize representations of statistically segmented words. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00447

Graf Estes, K., & Lew-Williams, C. (2015). Listening through voices: Infant statistical word segmentation across multiple speakers. *Developmental Psychology*, *51*(11), 1517–1528. https://doi.org/10.1037/a0039725

Gray, S. (2004). Word learning by preschoolers with specific language impairment: predictors and poor learners. *Journal of Speech, Language,* *and Hearing Research*, *47*(5), 1117–1132. https://doi.org/10.1044/1092-4388(2004/083)

Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders*, *39*(2), 158–170. https://doi.org/10.1016/j.jcomdis.2005.11.004

Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(11), 1251–1263. https://doi.org/10.1111/jcpp.12734

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018a). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018b). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115

Harnad, S. R. (1990). *Categorical perception: the groundwork of cognition*. Cambridge University Press.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*(1), 65–94. https://doi.org/10.1177/0267658307071601

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., Alm, P., Jennische, M., Bruce Tomblin, J., & Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with Specific Language Impairment. *Research in Developmental Disabilities*, *32*(6), 2362–2375. https://doi.org/10.1016/j.ridd.2011.07.026

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. https://doi.org/10.1080/15250000701795598

Hsu, H. J., & Bishop, D. V. (2010). Grammatical difficulties in children with specific language impairment: Is learning deficient? *Human Development*, *53*(5), 264–277. https://doi.org/10.1159/000321289

Hsu, H. J., & Bishop, D. V. (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental Science*, *17*(3), 352–365. https://doi.org/10.1111/desc.12125

Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00175

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition*, *225*, 105123. https://doi.org/10.1016/J.COGNITION.2022.105123

Jackson, E., Leitao, S., Claessen, M., & Boyes, M. (2019). Fast mapping short and long words: Examining the influence of phonological short-term memory and receptive vocabulary in children with developmental language disorder. *Journal of Communication Disorders*, *79*, 11–23. https://doi.org/10.1016/j.jcomdis.2019.02.001

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2019). The evaluation of word-learning abilities in people with developmental language disorder: a scoping review. *International Journal of Language and Communication Disorders*, *54*(5), 742–755. https://doi.org/10.1111/1460-6984.12490

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific Language Impairment: a deficit in grammar or processing? *Trends in Cognitive Sciences*, *2*(7), 240–247.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345. https://doi.org/10.1111/j.1467-7687.2009.00886.x

Junge, C., van Rooijen, R., & Raijmakers, M. (2018). Distributional information shapes infants' categorization of objects. *Infancy*, *23*(6), 917–926. https://doi.org/10.1111/infa.12258

Kachergis, G., Yu, C., & Shiffrin, R. M. (2014). Cross-situational word learning is both implicit and strategic. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00588

Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: a meta-analysis. *Journal of Speech, Language and Hearing Research*, *53*(3), 739–756.

Kapa, L. L., & Erikson, J. A. (2020). The relationship between word learning and executive function in preschoolers with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2293–2307. https://doi.org/10.1044/2020_JSLHR-19-00342

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321–340. https://doi.org/10.1016/j.cognition.2010.05.011

Kautto, A., & Mainela-Arnold, E. (2022). Procedural learning and school-age language outcomes in children with and without a history of late talking. *International Journal of Language & Communication Disorders*. https://doi.org/10.1111/1460-6984.12751

Kemény, F., & Lukács, Á. (2010). Impaired procedural learning in language impairment: Results from probabilistic categorization. *Journal of Clinical and Experimental Neuropsychology*, *32*(3), 249–258. https://doi.org/10.1080/13803390902971131

Kemény, F., & Lukács, Á. (2021). The role of statistical learning and verbal short-term memory in impaired and typical lexical development. *Frontiers in Communication*, *6*. https://doi.org/10.3389/fcomm.2021.700452

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184. https://doi.org/10.1037/a0025405

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184–193. https://doi.org/10.1111/cdev.12461

Kidd, E., Arciuli, J., Christiansen, M. H., Isbilen, E. S., Revius, K., & Smithson, M. (2020). Measuring children's auditory statistical learning via serial recall. *Journal of Experimental Child Psychology*, *200*. https://doi.org/10.1016/j.jecp.2020.104964

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, *24*(1), 109–116. https://doi.org/10.1016/j.cogbrainres.2004.12.009

Kraemer, H. C., & Blasey, C. M. (2004). Centring in regression analyses: a strategy to prevent errors in statistical inference. *International Journal of Methods in Psychiatric Research*, *13*(3), 141–151. https://doi.org/10.1002/mpr.170

Krzemien, M., Thibaut, J. P., Jemel, B., Levaux, E., & Maillart, C. (2021). How do children with developmental language disorder extend novel nouns? *Journal of Experimental Child Psychology*, *202*. https://doi.org/10.1016/J.JECP.2020.105010

Lahey, M., & Edwards, J. (1999). Naming errors of children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(1), 195–205. https://doi.org/10.1044/jslhr.4201.195

Lammertink, I. (2020). *Detecting patterns*. University of Amsterdam.

Lammertink, I., Boersma, P., Rispens, J., & Wijnen, F. (2020). Visual statistical learning in children with and without DLD and its relation to literacy in children with DLD. *Reading and Writing*, *33*(6), 1557–1589. https://doi.org/10.1007/s11145-020-10018-4

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: a meta-analysis. *Journal of Speech, Language and Hearing Research*, *60*(12), 1–13. https://doi.org/10.23641/asha

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2019). Children with developmental language disorder have an auditory verbal statistical learning deficit: evidence from an online measure. *Language Learning*, *70*(1), 137–148. https://doi.org/10.1111/lang.12373

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2020). Statistical learning in the visuomotor domain and its relation to grammatical proficiency in children with and without developmental language disorder: a conceptual replication and meta-analysis. *Language Learning and Development*, *16*(4), 426–450. https://doi.org/10.1080/15475441.2020.1820340

Leonard, L. B. (2014). *Children with specific language impairment*. MIT Press.

Leonard, L. B., Nippold, M. A., Kail, R., & Hale, C. A. (1983). Picture naming in language-impaired children. *Journal of Speech and Hearing Research*, *26*(4), 609–615. https://doi.org/10.1044/jshr.2604.609

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Lukács, Á., & Kemény, F. (2014). Domain-general sequence learning deficit in specific language impairment. *Neuropsychology*, *28*(3), 472–483. https://doi.org/10.1037/neu0000052

Lukács, Á., Lukics, K. S., & Dobó, D. (2021). Online statistical learning in developmental language disorder. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.715818

Lukics, K. S., & Lukács, Á. (2021). Tracking statistical learning online: Word segmentation in a target detection task. *Acta Psychologica*, *215*. https://doi.org/10.1016/j.actpsy.2021.103271

Lum, J. A. G., & Clark, G. M. (2022). Implicit manual and oculomotor sequence learning in developmental language disorder. *Developmental Science*, *25*(2). https://doi.org/10.1111/desc.13156

Lum, J. A. G., & Conti-Ramsden, G. (2013). Long-term memory: A review and meta-analysis of studies of declarative and procedural memory in specific language impairment. *Topics in Language Disorders*, *33*(4), 282–297. https://doi.org/10.1097/01.TLD.0000437939.01237.6a

Lum, J. A. G., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex*, *51*(1), 1–10. https://doi.org/10.1016/j.cortex.2013.10.011

Lum, J. A. G., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*(9), 1138–1154. https://doi.org/10.1016/j.cortex.2011.06.001

MacNamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review*, *79*(1), 1–13. https://doi.org/10.1037/h0031901

Mainela-Arnold, E., & Evans, J. L. (2014). Do statistical segmentation abilities predict lexical-phonological and lexical-semantic abilities in children with and without SLI? *Journal of Child Language*, *41*(2), 327–351. https://doi.org/10.1017/S0305000912000736

Mainela-Arnold, E., Evans, J. L., & Coady, J. A. (2010). Explaining lexical-semantic deficits in specific language impairment: the role of phonological similarity, phonological working memory, and lexical competition. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1742–1756. https://doi.org/10.1044/1092-4388(2010/08-0198)

Marinellie, S. A., & Johnson, C. J. (2002). Definitional skill in school-age children with specific language impairment. *Journal of Communication Disorders*, *35*(3), 241–259. https://doi.org/10.1016/S0021-9924(02)00056-4

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134. https://doi.org/10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(2), B101–B111. https://doi.org/10.1016/s0010-0277(01)00157-3

Mayor-Dubois, C., Zesiger, P., van der Linden, M., & Roulet-Perez, E. (2014). Nondeclarative learning in children with Specific Language Impairment: Predicting regularities in the visuomotor, phonological, and cognitive domains. *Child Neuropsychology*, *20*(1), 14–22. https://doi.org/10.1080/09297049.2012.734293

McGregor, K. K. (1997). The nature of word-finding errors of preschoolers with and without word- finding deficits. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1232–1244. https://doi.org/10.1044/jslhr.4006.1232

McGregor, K. K., & Appel, A. (2002). On the relation between mental representation and naming in a child with specific language impairment. *Clinical Linguistics and Phonetics*, *16*(1), 1–20. https://doi.org/10.1080/02699200110085034

McGregor, K. K., Arbisi-Kelm, T., & Eden, N. (2017). The encoding of word forms into memory may be challenging for college students with developmental language impairment. *International Journal of Speech-Language Pathology*, *19*(1), 43–57. https://doi.org/10.3109/17549507.2016.1159337

McGregor, K. K., Berns, A. J., Owen, A. J., Michels, S. A., Duff, D., Bahnsen, A. J., & Lloyd, M. (2012). Associations between syntax and the lexicon among children with or without ASD and language impairment. *Journal of Autism and Developmental Disorders*, *42*(1), 35–47. https://doi.org/10.1007/s10803-011-1210-4

McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *45*(5), 998–1014. https://doi.org/10.1044/1092-4388(2002/081)

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language and Communication Disorders*, *48*(3), 307–319. https://doi.org/10.1111/1460-6984.12008

McGregor, K. K., Smolak, E., Jones, M., Oleson, J., Eden, N., Arbisi-Kelm, T., & Pomper, R. (2022). What children with developmental language disorder teach us about cross-situational word learning. *Cognitive Science*, *46*(2). https://doi.org/10.1111/cogs.13094

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872.supp

Melby-Lervåg, M., Lervåg, A., Lyster, S. A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science*, *23*(10), 1092–1098. https://doi.org/10.1177/0956797612443833

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, *91*(1), 3–19. https://doi.org/10.1037/0022-0663.91.1.3

Micheyl, C., Hunter, C., & Oxenham, A. J. (2010). Auditory stream segregation and the perception of across-frequency synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 1029–1039. https://doi.org/10.1037/a0017601

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *44*(2), 416–433. https://doi.org/10.1044/1092-4388(2001/034

Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*(1), 271–280. https://doi.org/10.1016/j.cognition.2008.02.003

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00031

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and specific language impairment: an update on the relation and perspectives on assessment and treatment. *American Journal of*

Speech-Language Pathology, *19*(1), 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028)

Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research*, *48*(2), 439–458. https://doi.org/https://doi.org/10.1044/1092-4388(2005/030)

Nation, K. (2014). Lexical learning and lexical processing in children with developmental language impairments. *Philosophical Transactions of the Royal Society B*, *369*(1634). https://doi.org/10.1098/rstb.2012.0387

Newell, F. N., & Bülthoff, H. H. (2002). Categorical perception of familiar objects. *Cognition*, *85*(2), 113–143. https://doi.org/10.1016/S0010-0277(02)00104-X

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, *42*(4), 643–655. https://doi.org/10.1037/0012-1649.42.4.643

Ng, C. S. Y., Stokes, S. F., & Alt, M. (2020). Successful implicit vocabulary intervention for three Cantonese-speaking toddlers: a replicated single-case design. *Journal of Speech, Language, and Hearing Research*, *63*(12), 4148–4161. https://doi.org/10.1044/2020_JSLHR-20-00087

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience 2011 14:9*, *14*(9), 1105–1107. https://doi.org/10.1038/nn.2886

Noonan, N. B. (2018). *Exploring the process of statistical language learning*. The University of Western Ontario.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory. In *McGraw-Hill, New York* (Issue 1994). McGraw-Hill.

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. G. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology*, *7*(\). https://doi.org/10.3389/fpsyg.2016.01245

Paciorek, A., & Williams, J. N. (2015). Semantic Implicit learning. In *Implicit and explicit learning of languages* (pp. 69–90). John Benjamins.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, *5*(4), 291–303. https://doi.org/10.1038/nrn1364

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: exemplar variability supports higher-order generalization. *Psychological Science*, *21*(12), 1894–1902. https://doi.org/10.1177/0956797610389189

Pinker, S. (1989). Resolving a learnability paradox in the acquisition of the verb lexicon. In M. L. Rice & R. L. Schiefelbusch (Eds.), *The teachability of language* (pp. 13–61). Paul H. Brooks Publishing.

Plante, E., Gomez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, *35*(5), 453–462. https://doi.org/10.1016/S0021-9924(02)00094-1

Plante, E., & Gómez, R. L. (2018). Learning without trying: the clinical relevance of statistical learning. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 710–722. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0131

Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., Lieser, A. M., & Burton, R. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology*, *23*(4), 530–545. https://doi.org/10.1044/2014_AJSLP-13-0038

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–681. https://doi.org/10.1016/j.cognition.2007.04.003

Psychology Software Tools Inc. (2002). *E-Prime* (2.0). https://support.pstnet.com/

Psychology Software Tools Inc. (2016). *E-Prime* (3.0). https://support.pstnet.com/

Quam, C., Cardinal, H., Gallegos, C., & Bodner, T. (2020). Sound discrimination and explicit mapping of sounds to meanings in preschoolers with and without developmental language disorder.

International Journal of Speech-Language Pathology, *23*(1), 26–37. https://doi.org/10.1080/17549507.2020.1750701

Quine, W. V. O. (1960). *Word and object*. MIT Press.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ralli, A. M., Dockrell, J., & Dockrell, J. E. (2010). Real world word learning: exploring the development of children's lexical representations. *Psychology*, *17*(1), 1–24. https://doi.org/10.12681/psy_hps.23750

Raven, J., Raven, J., & Court, J. (2003). *Manual for Raven's progressive matrices and vocabulary scales*. Harcourt Assessment.

Rice, M. L., Buhr, J. C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders*, *55*(1), 33–42. https://doi.org/10.1044/jshd.5501.33

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*, *37*(1), 106–122. https://doi.org/10.1044/jshr.3701.106

Rice, M., Wexler, K., & Cleave, P. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research*, *38*(4), 850–863. https://doi.org/10.1044/jshr.3804.850

Rispens, J., & Baker, A. (2012). Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research*, *55*(3), 683–694. https://doi.org/10.1044/1092-4388(2011/10-0263)

Roembke, T. C., Mcmurray, B., Author, C., & Jaegerstrasse, T. R. (2020). *Interactions of explicit and implicit learning mechanisms in cross-situational word learning*.

Rogers, L. L., Park, S. H., & Vickery, T. J. (2021). Visual statistical learning is modulated by arbitrary and natural categories. *Psychonomic Bulletin & Review*, *28*(4), 1281–1288. https://doi.org/10.3758/s13423-021-01917-w

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Sack, L., Dollaghan, C., & Goffman, L. (2021). Contributions of early motor deficits in predicting language outcomes among preschoolers with developmental language disorder. *International Journal of Speech-Language Pathology*, *24*(4), 1–13. https://doi.org/10.1080/17549507.2021.1998629

Saffran, J. R. (2018). Statistical learning as a window into developmental disabilities. *Journal of Neurodevelopmental Disorders*, *10*(1). https://doi.org/10.1186/s11689-018-9252-y

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Sandgren, O., Salameh, E. K., Nettelbladt, U., Dahlgren-Sandberg, A., & Andersson, K. (2021). Using a word association task to investigate semantic depth in Swedish-speaking children with developmental language disorder. *Logopedics Phoniatrics Vocology*, *46*(3), 134–140. https://doi.org/10.1080/14015439.2020.1785001

Sanjeevan, T., & Mainela-Arnold, E. (2019). Characterizing the motor skills in children with specific language impairment. *Folia Phoniatrica et Logopaedica*, *71*(1), 42–55. https://doi.org/10.1159/000493262

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

Semel, E., Wiig, E., & Secord, W. (2010). *Clinical evalutation of language fundamentals: Dutch version* (W. Kort, E. Compaan, M. Schittekatte, & P. Dekker, Eds.; Third edition). Pearson.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: domain-general and domain-specific associations with language. *Infancy*, *17*(3), 247–271. https://doi.org/10.1111/j.1532-7078.2011.00085.x

Sheng, L., & McGregor, K. K. (2010). Lexical-semantic organization in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *53*(1), 146–159. https://doi.org/10.1044/1092-4388(2009/08-0160)

Sherman, B. E., Graves, K. N., & Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Current Opinion in Behavioral Sciences*, *32*, 15–20. https://doi.org/10.1016/j.cobeha.2020.01.015

Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass*, *14*(3). https://doi.org/10.1111/lnc3.12365

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0059

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432. https://doi.org/10.3758/s13428-016-0719-z

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: what does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, *42*, 692–727. https://doi.org/10.1111/cogs.12556

Singh, L., Steven Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, *15*(4), 482–495. https://doi.org/10.1111/j.1467-7687.2012.01141.x

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge University Press.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*(6), 246–251. https://doi.org/10.1016/S1364-6613(03)00109-8

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480–498. https://doi.org/10.1111/j.1551-6709.2010.01158.x

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Smolak, E., McGregor, K. K., Arbisi-Kelm, T., & Eden, N. (2020). Sustained attention in developmental language disorder and its relation to working memory and language. *Journal of Speech, Language, and Hearing Research*, *63*(12), 4096–4108. https://doi.org/10.1044/2020_JSLHR-20-00265

Snowling, M. J., Hayiou-Thomas, M. E., Nash, H. M., & Hulme, C. (2020). Dyslexia and developmental language disorder: comorbid disorders with distinct effects on reading comprehension. *Journal of Child Psychology and Psychiatry*, *61*(6), 672–680. https://doi.org/10.1111/JCPP.13140

Sociaal en Cultureel Planbureau. (2018). https://www.scp.nl/

Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and Writing*, *28*(4), 467–490. https://doi.org/10.1007/s11145-014-9533-0

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Suanda, S. H., & Namy, L. L. (2012). Detailed behavioral analysis as a window into cross-situational word learning. *Cognitive Science*, *36*(3), 545–559. https://doi.org/10.1111/j.1551-6709.2011.01218.x

Tallal, P. (1990). Fine-grained discrimination deficits in language-learning impaired children are specific neither to the auditory modality nor to speech perception. *Journal of Speech and Hearing Research*, *33*(3), 617–618. https://doi.org/10.1044/JSHR.3303.616

ter Schure, S. M. M., Junge, C. M. M., & Boersma, P. P. G. (2016). Semantics guide infants' vowel learning: Computational and experimental evidence. *Infant Behavior and Development*, *43*, 44–57. https://doi.org/10.1016/j.infbeh.2016.01.002

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Tobii Pro AB. (2014). *Tobii Pro Lab user's manual*. Tobii Pro.

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, *3*(4), 269–293. https://doi.org/10.1080/15475440701377477

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2). https://doi.org/10.1016/j.cognition.2005.01.006

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Ullman, M. T. (2016). The declarative/procedural model: a neurobiological model of language learning, knowledge, and use. In *Neurobiology of Language* (pp. 953–968). Elsevier Inc. https://doi.org/10.1016/B978-0-12-407794-2.00076-6

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex*, *41*(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience &*

Biobehavioral Reviews, *51*, 205–222. https://doi.org/10.1016/J.NEUBIOREV.2015.01.008

Unger, L., & Fisher, A. v. (2021). The emergence of richly organized semantic knowledge from simple statistics: A synthetic review. *Developmental Review*, *60*. https://doi.org/10.1016/j.dr.2021.100949

van Daal, J., Verhoeven, L., & van Balkom, H. (2004). Subtypes of severe speech and language impairments: psychometric evidence from 4-year-old children in the Netherlands. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1411–1423. https://doi.org/10.1044/1092-4388(2004/105

van den Bedem, N. P., Alphen, V., Rooij, D., Samson, M., & Harjunen, A. C. (2019). Depression in children with developmental language disorders. *Word of Mouth*, *30*(4), 1–4.

van den Bedem, N. P., Dockrell, J. E., van Alphen, P. M., Kalicharan, S. v., & Rieffe, C. (2018). Victimization, bullying, and emotional competence: longitudinal associations in (pre)adolescents with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *61*(8), 2028–2044. https://doi.org/10.1044/2018_JSLHR-L-17-0429

van der Lely, H. K. J. (2005). Domain-specific cognitive systems: insight from Grammatical-SLI. *Trends in Cognitive Sciences*, *9*(2), 53–59. https://doi.org/10.1016/J.TICS.2004.12.002

van Hedger, S. C., Johnsrude, I. S., & Batterink, L. J. (2022). Musical instrument familiarity affects statistical learning of tone sequences. *Cognition*, *218*, 104949. https://doi.org/10.1016/J.COGNITION.2021.104949

van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences. In *IPO Annual Progress Report* (Vol. 9). Technische Hogeschool Eindhoven.

Vandermosten, M., Wouters, J., Ghesquière, P., & Golestani, N. (2019). Statistical learning of speech sounds in dyslexic and typical reading children. *Scientific Studies of Reading*, *23*(1), 116–127. https://doi.org/10.1080/10888438.2018.1473404

Venker, C. E. (2019). Cross-situational and ostensive word learning in children with and without autism spectrum disorder. *Cognition*, *183*, 181–191. https://doi.org/10.1016/j.cognition.2018.10.025

Viviani, P. (1990). Eye movement in visual search: cognitive, perceptual and motor control aspects. In E. Kowler (Ed.), *Eye movements and their role in visual cognitive processes* (pp. 353–393).

Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of Memory and Language*, *93*, 217–230. https://doi.org/10.1016/j.jml.2016.10.001

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382. https://doi.org/10.1016/j.cognition.2013.02.015

von Koss Torkildsen, J., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *Journal of Speech, Language, and Hearing Research*, *56*(2), 618–629. https://doi.org/10.1044/1092-4388(2012/11-0125)

Wanrooij, K., Boersma, P., & Benders, T. (2015). Observed effects of "distributional learning" may not relate to the number of peaks. A test of "dispersion" as a confounding factor. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01341

Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00077

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

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. https://doi.org/10.1080/10888438.2020.1750618

Witzel, C., & Gegenfurtner, K. R. (2016). Categorical perception for red and brown. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 540–570. https://doi.org/10.1037/xhp0000154

Wu, R., Gopnik, A., & Richardson, D. (2010). Social cues support learning about objects from statistics in infancy. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*, 1228–1233. https://doi.org/10.1037/a0024023

Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, *47*(5), 1220–1229. https://doi.org/10.1037/a0024023

Yang, Y., Yim, D., Kim, S., & Han, J. (2013). The relationship among receptive vocabulary, non-word repetition, and quick incidental learning in preschoolers with and without delay in vocabulary development. *Communication Sciences and Disorders*, *18*(4), 379–391. https://doi.org/10.12963/CSD.13079

Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, *56*(6), 1574–1583. https://doi.org/10.2307/1130476

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: integrating statistical and social cues. *Neurocomputing*, *70*(13–15), 2149–2165. https://doi.org/10.1016/j.neucom.2006.01.034

Yu, C., & Smith, L. (2012). Modeling cross-situational word–referent learning: prior questions. *Psychological Review*, *119*(1), 21–39. https://doi.org/10.1037/a0026182.supp

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180. https://doi.org/10.1111/j.1467-7687.2010.00958.x

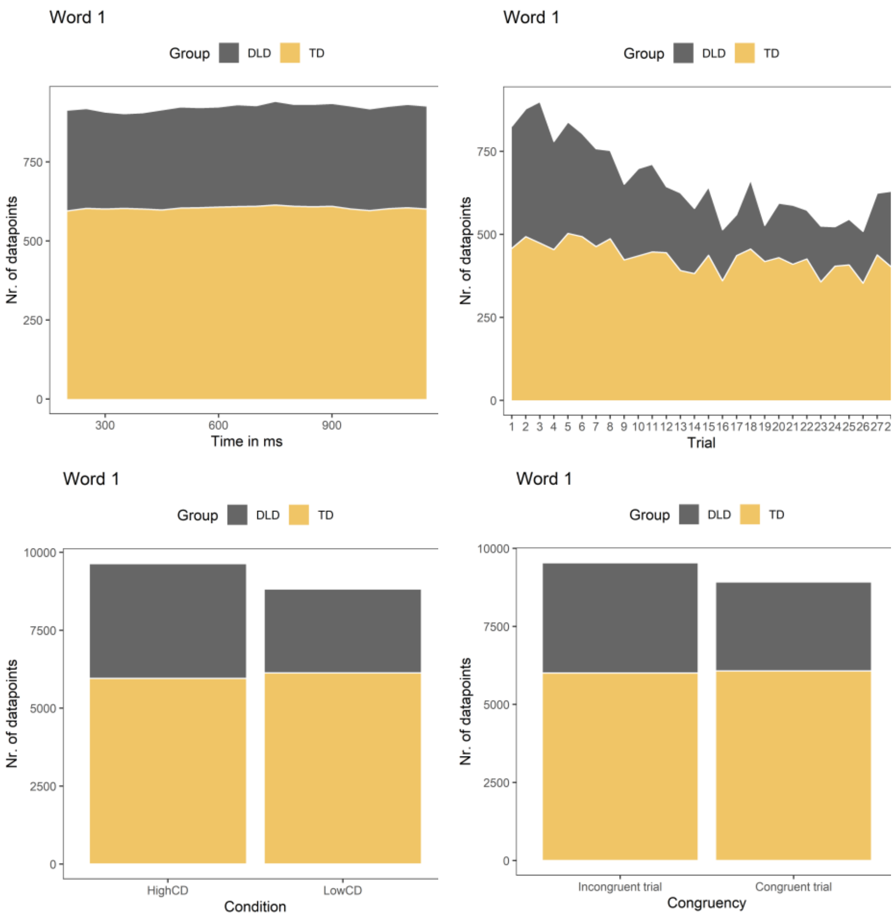
Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross-situational statistical learning: Evidence from eye tracking. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00148

Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin and Review*, *21*(1), 1–22. https://doi.org/10.3758/s13423-013-0443-y

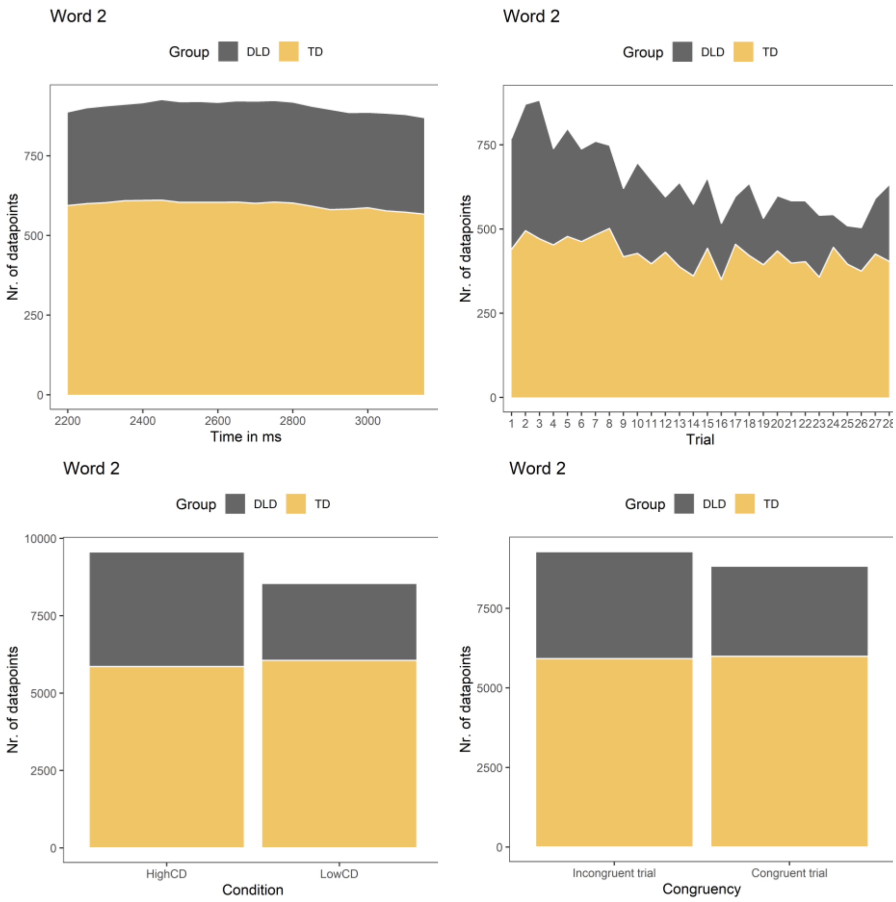
Zhang, Y., Yurovsky, D., & Yu, C. (2021). Cross-situational learning from ambiguous egocentric input is a continuous process: evidence using the human simulation paradigm. *Cognitive Science*, *45*(7). https://doi.org/10.1111/cogs.13010

Zsef Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826. https://doi.org/10.1073pnas.232472899

Appendix



Appendix 1 – Number of data points eye-tracking data for Word1 (chapter 3).



Appendix 2 – Number of data points eye-tracking data for Word2 (chapter 3).

Summary

Lexical-semantic deficits in developmental language disorder: the role of statistical learning

Children with developmental language disorder (DLD) have serious difficulties with learning to speak and understand language, which is, when you think about it, not an easy task at all. Language is full of patterns and regularities. For example, certain sounds occur together frequently in Dutch (*tr-*) while others do not (*tl-*). Children have to learn about all these patterns to become proficient speakers of their native language(s). That children can do this, could be thanks to a learning mechanism called "statistical learning". Statistical learning entails the implicit learning of patterns in all kinds of input (verbal, visual, auditory, and so on). For example, a person that is subjected to a stream of syllables (…*bidakutupirogolabutupirobidakugolabu*…) for a few minutes, usually learns that some syllables (such as *tupiro*) occur together more frequently than other combinations of syllables (such as *rogola*), while they might not even be aware of this newly developed knowledge. If this learning mechanism is indeed important for language acquisition, could it be the case that children that have difficulty with language acquisition (children with DLD) have a deficit in this type of learning? This dissertation focuses on that question.

Children with DLD can have difficulties in all levels of language, but this dissertation focuses on their problems with the development of lexical-semantic knowledge in particular. Compared to children without DLD (typically developing; TD) children, children with DLD know fewer words, and their word knowledge is more superficial. For example, they might have difficulty with providing definitions of common words, or even draw less detailed pictures of concepts compared to TD children. The network of words in their mental lexicon also seems to be organized less efficiently, which could cause trouble with finding the right words quickly. Thus, learning all semantic (meaning-related) aspects of words in the lexicon and learning how to use words correctly and fluently in the right context, can be very difficult for children with DLD. Unsurprisingly, these difficulties could strongly affect academic and social development. In this dissertation, we aimed to test the hypothesis that children with DLD have a deficit in statistical learning and that this deficit contributes to their lexical-semantic difficulties. For this end, we developed three tasks that targeted statistical learning of various types of lexical-semantic knowledge: the word segmentation task (Chapter 2), the cross-situational word learning task (Chapter 3) and the visual distributional learning task (Chapter 4 and Chapter 5). The tasks and our found results are explained in more detail below. We expected that children with DLD (7-9 years old) show a deficit compared to TD children in all three types of statistical learning. Moreover, we expected to detect correlations between performance on the tasks and measures of lexical-semantic ability; in the sense that children who are less proficient in statistical learning also have weaker lexical-semantic skills.

Our first task, the word segmentation task, is designed to mimic the process of finding word boundaries in fluent speech. In actual speech, boundaries between words are not consistently marked by pauses or other prosodic cues. One can experience this when listening to an unknown language: it seems impossible to know where words start and end. Infants face the task of finding words in the new language they are immersed in. In Chapter 2 we report our version of the word segmentation task. During the task, participants were subjected to an uninterrupted stream of syllables for eight minutes. This stream consisted of the repetition of four "words" (*kiba*, *moti*, *dalu* and *gido*). The words were recorded by a female native speaker of Dutch, but edited into monotone speech. All syllables were equally long and there were no pauses between them. The words were repeated in a randomized order, with the restriction that words could not be repeated twice in a row. The stream could thus sound like: …*kibadalumotikibadalugidomoti*… The probability of syllables following up on each other was different within words than between words. For example, the probability that *ba* followed *ki* was 100%, while the probability that *da* followed *ba* was 33%, as *kiba* was a "word", but *ba-da* was not.

After the exposure phase, participants answered 16 test questions. During every question, the participant heard two sequences of syllables: one (the target) corresponded to one of the "words", for example *kiba*. The other sequence (the "foil") was a combination of syllables that had sometimes occurred in the stream, but only 33% of the time (for example *ba-da*, as *dalu* only followed *kiba* 33% of the time). We expected that participants would select the targets more often than the foils, indicating that people become sensitive to the word boundaries in the stream. However, in our first group of adult participants, we were not able to show that performance was better than would be expected by chance; in other words, participants did not seem to have developed a preference for either the targets or the foils. After this first group of participants, we tested two other groups of adults and one group of school-aged children, and made modifications to the task. However, in none of the tested groups we found a significant learning effect. As our task did not seem to be a reliable measure of this type of statistical learning, we unfortunately were not able to compare word segmentation ability in TD children and children with DLD.

Besides discovering words in fluent speech, children need to learn the meaning of those words. For example, they need to match the word *goat* to the referent GOAT in their environment. This task can be quite difficult when you consider the vast number of words and potential referents in a child's environment at a given moment in time. Many word-learning situations are ambiguous as it is not always clear which word refers to what referent. Research has shown that through statistical learning, people can learn these mappings across situations: a process called cross-situational word learning. While a given situation can be ambiguous in itself (for example: the words *goat* and *cow* are uttered, amongst other words, while the child sees various farm animals in the visual environment), the correct word–referent mappings can be learned *across* situations, as a word and its corresponding referent co-occur frequently in different situations. In Chapter 3 we report the performance of children with and without DLD on a cross-situational word learning task. The children were subjected to an exposure phase with learning trials that consisted of two unknown objects and two novel words, without indication of which word referred to which referent. Across trials, a word and its referent co-occurred consistently. Eight word–referent pairs were presented on 28 exposure trials. No explicit instructions were given: the children were only told to pay attention to the pictures and words, and that there would be questions afterwards. During those questions, all words were played once, and the children were asked to choose which object corresponded to it. Our results show that both children with and without DLD are able to learn word–referent pairs in an implicit cross-situational word learning task, but that TD children are more proficient than children with DLD. Thus, children with DLD might need more exposure to reach the same level as TD children.

Another stage in word learning is the formation of semantic categories. For example, children need to learn the differences and similarities between cows and goats to form the categories COW and GOAT. Distributional learning is a type of statistical learning that is important for categorization of auditory and visual stimuli. In Chapter 4 we report visual distributional learning in TD school-aged children, while we compare this type of learning between TD children and children with DLD in Chapter 5. In the task, children were exposed to novel visual stimuli, which formed a continuum from one endpoint stimulus to the other endpoint stimulus in ten equal steps (see Figure 1). During exposure, some tokens were shown more frequently than others. There were two conditions that differed in the frequency distribution of the tokens in the exposure phase, indicating different underlying categories, as the 'frequency peaks' occurred at different positions in the continuum. After exposure, it was tested how children in the different conditions categorized the tokens. Overall, children indeed show sensitivity to distributional properties of visual input, as the exposure condition significantly influenced categorization. We did not find evidence for a difference between children with and without DLD in this respect; we thus cannot say whether children with DLD have a visual distributional learning deficit.

Figure 1 – Stimuli used for the visual distributional learning task.

Besides testing whether children with DLD have a deficit in various types of statistical learning, we also wanted to know whether their statistical learning ability correlates with their lexical-semantic knowledge. The children with DLD were tested with various standardized measures of lexical-semantic ability (receptive vocabulary size, productive vocabulary size, word category knowledge and lexical-semantic organization), as well as various control measures (non-verbal intelligence, working memory, phonological processing). In Chapter 3, we investigated whether implicit cross-situational word learning ability significantly correlated with measures of lexical-semantic knowledge in children with DLD, while taking variation in the control measures into account. In Chapter 5, we did the same for visual distributional learning ability. Unfortunately, we did not find any evidence for or against these relationships. We thus have inconclusive results and cannot say whether statistical learning ability contributes to lexical-semantic knowledge in children with DLD. It could be the case that our statistical learning tasks (and statistical learning tasks in general) are not reliable measures of statistical learning ability at the individual level, as they are designed for group-level comparisons.

The final aim of this dissertation was to find "on-line measures" of statistical learning. Measuring statistical learning on-line entails measuring learning already during exposure to the input, instead of only *after* learning is supposed to have taken place. On-line measures have the potential to reveal more fine-grained information about the process of statistical learning in different groups of participants. Moreover, "off-line measures" of statistical learning, such as the test questions described in the word segmentation task in Chapter 2, are often difficult for children and likely influenced by memory and attention abilities. Therefore we added on-line measures of learning to the word segmentation task (Chapter 2) and the cross-situational word learning task (Chapter 3). In the word segmentation task, we aimed to measure learning with the use of a click detection task. Click sounds were added to the stream of syllables in different positions: either between the “words” (*kiba!dalu*), or within a “word” (*ki!ba*). Participants were prompted to push a button as fast as they could when they heard a click sound. We hypothesized that if participants had gained some knowledge of the word boundaries in the stream, clicks within words would be more unexpected (after *ki*, a participant would expect *ba* and not a click sound) than when a click was placed between two words, therefore slowing down reaction times to clicks within words compared to clicks between words. Our expectations were not met at the group level. Some participants indeed showed the expected effect, but others showed an effect in the opposite direction. During the cross-situational word learning task, we measured our participants’ eye gaze during the exposure phase as an on-line measure of statistical learning. We expected that as participants gained knowledge about the word–referent pairs, they would start looking more towards the correct image than the distractor image. Our children indeed looked significantly more towards the target image overall, but there was no significant change during the course of the experiment nor a significant difference between our children with DLD and our TD children. We thus cannot conclude whether there is a difference in the learning trajectory between the two groups of children.

In Chapter 6 we discuss the results from our different studies. We conclude that children with DLD likely have more difficulty than TD children with using statistical learning mechanisms to map words to referents. This could hamper word learning and cause lexical-semantic difficulties – although we cannot underpin this hypothesis with evidence for a significant relationship between implicit cross-situational word learning and lexical-semantic knowledge in children with DLD. We have no evidence for a deficit in visual distributional learning. This does not necessarily mean that the statistical learning deficit in children with DLD is confined to verbal statistical learning. Systematic comparisons between performance on statistical learning tasks in different domains, modalities and with different underlying statistical regularities should be made to find out the scope of the statistical learning deficit in children with DLD. Our results concerning the relationship between statistical learning and lexical-semantic knowledge in children with DLD are inconclusive. Future research could aim test larger groups of participants, conduct more ecologically valid statistical learning tasks and find ways to measure statistical learning ability on the individual level more reliably.

Samenvatting

Kinderen met een taalontwikkelingsstoornis hebben moeite met woordbetekenis: de rol van statistisch leren

Kinderen met een taalontwikkelingsstoornis (TOS) hebben veel moeite met het gebruiken en begrijpen van taal. Waar gaat het mis? Taal zit vol met terugkerende patronen. In het Nederlands bijvoorbeeld, komen binnen lettergrepen bepaalde klankcombinaties vaak voor (zoals *tr-*), en andere bijna nooit (zoals *tl-*). Kinderen moeten heel veel van dit soort patronen leren om een vaardige spreker van hun moedertaal te worden. Dat kinderen dit kunnen, is wellicht dankzij een leermechanisme dat "statistisch leren" genoemd wordt. Statistisch leren gaat over het impliciet leren van patronen in verschillende soorten input (verbaal, visueel, auditief, etc.). Iemand die bijvoorbeeld een paar minuten lang wordt blootgesteld aan een stroom van lettergrepen (…*bidakutupirogolabutupirobidakugolabu*…), zal waarschijnlijk leren dat sommige lettergrepen (zoals *tupiro*) vaker samen voorkomen dan andere combinaties van lettergrepen (zoals *rogola*), terwijl diegene zich misschien niet eens bewust is van deze opgedane kennis. Als dit leermechanisme inderdaad belangrijk is voor taalverwerving, zou het dan kunnen dat kinderen die moeite hebben met de verwerving van taal, moeilijkheden hebben met dit type leren? Deze dissertatie richt zich op die onderzoeksvraag.

Kinderen met een TOS kunnen problemen hebben in alle taalgebieden, maar deze dissertatie richt zich op hun problemen met de ontwikkeling van kennis over woordbetekenissen (lexicaal-semantische kennis). Kinderen met een TOS kennen namelijk gemiddeld minder woorden, en hun woordkennis is oppervlakkiger vergeleken met kinderen met een typische taalontwikkeling (TT). Ze vinden het bijvoorbeeld moeilijk om definities te geven van veelvoorkomende woorden, en maken minder gespecificeerde tekeningen van concepten vergeleken met kinderen met een TT. Het netwerk van woorden in hun mentale lexicon lijkt ook minder efficiënt te zijn georganiseerd, waardoor het vinden van het juiste woord minder snel en accuraat verloopt. Kortom, het leren van semantische (betekenis) aspecten van woorden in het lexicon en leren hoe woorden correct en vloeiend in de juiste context gebruikt worden, kan heel moeilijk zijn voor kinderen met een TOS. Dit kan uiteraard verstrekkende gevolgen hebben voor hun schoolprestaties en sociaal-emotionele ontwikkeling. In deze dissertatie testen wij de hypothese dat kinderen met een TOS een stoornis hebben in het statistischleervermogen, en dat (onder andere) deze stoornis zorgt voor hun lexicaal-semantische problemen. Om dit te testen hebben wij drie taken ontwikkeld die het statistisch leren van verschillende soorten lexicaal-semantische informatie nabootsen: een woordsegmentatietaak (hoofdstuk 2), een "cross-situational word learning" taak (hoofdstuk 3) en een visueeldistributioneelleertaak (hoofdstuk 4 en 5). De taken en de resultaten die we hebben gevonden, worden hieronder uitvoerig besproken. Onze verwachting was dat kinderen met een TOS (tussen de 7 en 9 jaar oud) meer moeite hebben met de statistischleertaken dan de kinderen met een TT. Ook verwachtten we correlaties te vinden tussen statistischleervermogen en lexicaal-semantische vaardigheden, in de zin dat kinderen die minder goed zijn in statistisch leren ook zwakkere lexicaal-semantische vaardigheden hebben.

De eerste taak, de woordsegmentatietaak, is ontworpen als nabootsing van het vinden van woordgrenzen in vloeiende spraak. In het dagelijks leven worden woordgrenzen niet consistent gemarkeerd door pauzes of andere prosodische cues. Dit merk je als je een voor jou onbekende taal hoort: het lijkt onmogelijk om te horen waar woorden beginnen en eindigen. Baby's worden ook ondergedompeld in een voor hun nog onbekende taal en moeten leren woorden van elkaar te onderscheiden. In hoofdstuk 2 rapporteren we onze woordsegmentatietaak. Tijdens de taak werden proefpersonen acht minuten lang blootgesteld aan een ononderbroken stroom van lettergrepen. De stroom bestond uit de herhaling van vier "woorden" (*kiba*, *moti*, *dalu* en *gido*). De woorden waren ingesproken door een moedertaalspreekster van het Nederlands, maar bewerkt tot monotone spraak: alle lettergrepen waren even lang, er waren geen pauzes en er was geen intonatie. De vier woorden werden herhaald in willekeurige volgorde, met de restrictie dat een woord niet twee keer achter elkaar kon voorkomen. Het kon dus klinken als: …*kibadalumotikibadalugidomoti*… De waarschijnlijkheid dat lettergrepen achter elkaar voorkwamen was anders binnen woorden dan tussen woorden. Bijvoorbeeld: de kans dat *ba* na *ki* kwam was 100%, terwijl de kans dat *da* na *ba* kwam 33% was, aangezien *kiba* een "woord" was maar *ba-da* niet.

Na blootstelling aan de stimuli kregen de participanten 16 testvragen. Bij elke vraag werden twee items afgespeeld: één daarvan (de target) kwam overeen met een van de "woorden", bijvoorbeeld *kiba*. Het andere item (de afleider) was een combinatie van lettergrepen die wel voorkwam in de stroom, maar slechts 33% van de tijd (bijvoorbeeld *ba-da*, want die lettergrepen kwamen alleen na elkaar voor als *dalu* na *kiba* volgde). We verwachtten dat de proefpersonen een voorkeur zouden hebben voor de targets, wat erop zou wijzen dat ze gevoelig waren geworden voor de woordgrenzen in de stroom. Onze eerste groep volwassen participanten lieten echter geen voorkeur voor de target of de afleider. Na deze groep deelnemers hebben we nog twee andere groepen volwassenen en een groep schoolgaande kinderen getest op verschillende versies van de taak, maar in geen van de groepen hebben we een significant leereffect gevonden, dat wil zeggen in geen van deze resultaten hadden de proefpersonen een voorkeur voor de target. Aangezien onze taak niet betrouwbaar lijkt te zijn om dit specifieke statistischleervermogen te meten, konden we helaas woordsegmentatievaardigheid niet vergelijken tussen kinderen met een TOS en kinderen met een TT.

Naast het ontdekken van woorden in een brei van spraakklanken, moeten kinderen ook de betekenis van die woorden leren. Ze moeten bijvoorbeeld het woord *geit* aan de referent GEIT in hun omgeving koppelen. Deze taak is moeilijker dan het lijkt. Als een kind een woord hoort zoals *geit*, zijn er vaak meerdere potentiële referenten aanwezig in de omgeving van dat kind (bijvoorbeeld ook een schaap, koe of een kip in de kinderboerderij). Veel momenten waarop een nieuwe woordbetekenis zou kunnen worden geleerd zijn ambigu, aangezien het niet altijd duidelijk is welk woord waaraan refereert. Onderzoek laat zien dat mensen door middel van statistisch leren de koppelingen tussen woorden en referenten kunnen leren, na ze te hebben gezien in verschillende contexten: een proces dat "cross-situational word learning" wordt genoemd. Een bepaalde situatie kan op zichzelf ambigu zijn (als bijvoorbeeld de woorden *geit* en *koe* kort na elkaar worden geuit, samen met andere woorden, terwijl het kind verschillende soorten boerderijdieren ziet), maar omdat een woord en zijn referent vaak samen voorkomen in verschillende contexten, kan de koppeling over de tijd heen gemaakt worden. In hoofdstuk 3 rapporteren we hoe kinderen met en zonder TOS presteren op een cross-situational word learning taak. De kinderen werden blootgesteld aan een leerfase waar ze in elke trial twee plaatjes van onbekende objecten zagen en twee onbestaande woorden hoorden. Er werd niet aangegeven welk plaatje bij welk woord hoorde, maar de woorden en hun bijbehorende referent kwamen wel consistent samen voor in de hele leerfase. Acht verschillende woord–referentparen werden gepresenteerd in 28 leertrials. Er waren geen expliciete instructies: de kinderen kregen alleen te horen dat ze goed op de plaatjes en woorden moesten letten, en dat er aan het eind van het experiment vragen zouden komen. Tijdens die vragen werden alle woorden één keer afgespeeld en moesten de kinderen kiezen welk plaatje bij het woord hoorde. Onze resultaten laten zien dat zowel kinderen met en zonder TOS woord–referentparen kunnen leren in een impliciete cross-situational word learning taak, maar kinderen met een TT zijn hier beter in. Het zou dus zo kunnen zijn dat kinderen met een TOS meer input nodig hebben om hetzelfde niveau te halen.

Een ander stadium in het woordleerproces is het vormen van semantische categorieën. Kinderen moeten bijvoorbeeld de verschillen en overeenkomsten tussen koeien en geiten leren om de categorieën KOE en GEIT te vormen. Distributioneel leren is een vorm van statistisch leren die belangrijk is voor het categoriseren van auditieve en visuele stimuli. In hoofdstuk 4 rapporteren we visueel distributioneel leren in schoolgaande kinderen met een TT, en dit type leren vergeleken we tussen kinderen met en zonder TOS in hoofdstuk 5. Tijdens de taak werden de kinderen blootgesteld aan nieuwe plaatjes, die samen een continuüm vormen van één eindpunt naar een ander eindpunt in tien gelijke stappen (zie Afbeelding 1). Tijdens de leerfase werden sommige tokens vaker getoond dan anderen. Er waren twee condities die verschilden in de frequentieverdeling van de tokens in de leerfase. Deze condities reflecteerden een verschil in onderliggende categorieën, aangezien de 'frequentiepieken' op verschillende plekken in het continuüm zaten. Na de leerfase werd getest hoe de kinderen in de twee verschillende condities de tokens van het continuüm categoriseerden. Over het algemeen blijken kinderen inderdaad gevoelig te zijn voor de distributionele informatie in visuele input, aangezien de conditie van de leerfase een significante invloed had op categorisatie tijdens de test. We vonden geen bewijs voor een verschil tussen kinderen met en zonder TOS in dit opzicht; we kunnen dus niet concluderen of kinderen met een TOS een stoornis in visueel distributioneel leren hebben.

Afbeelding 1 – Stimuli van de visueeldistributioneelleertaak.

Naast het onderzoeken of kinderen met een TOS een stoornis hebben in verschillende soorten statistisch leren, wilden we ook weten of hun vaardigheid in statistisch leren samenhangt met hun lexicaal-semantische kennis. De kinderen met een TOS zijn getest met verschillende gestandaardiseerde taken die lexicaal-semantische vaardigheid meten (receptieve en productieve woordenschat, kennis van woordcategorieën en lexicaal-semantische organisatie), alsmede een aantal controlematen (non-verbale intelligentie, werkgeheugen, fonologische verwerking). In hoofdstuk 3 hebben we onderzocht of de vaardigheid op een impliciete cross-situational word learning taak samenhangt met lexicaal-semantische kennis in kinderen met een TOS, rekening houdend met variatie in de controlematen. In hoofdstuk 5 deden we hetzelfde voor de vaardigheid op de visueeldistributioneelleertaak. We hebben geen bewijs voor of tegen deze verbanden kunnen vinden. Omdat onze resultaten niet eenduidig zijn, kunnen we niet concluderen of statistischleervermogen bijdraagt aan lexicaal-semantische kennis in kinderen met een TOS. Mogelijk zijn onze taken (en taken die statistisch leren meten in het algemeen) geen betrouwbare maat van statistischleervermogen in het individu, aangezien ze zijn ontworpen voor vergelijkingen op groepsniveau.

Het laatste doel van deze dissertatie was het vinden van "on-line maten" van statistisch leren. Dit houdt in dat al tijdens de blootstelling aan de input wordt gemeten of proefpersonen leren, in plaats van dit alleen te meten *nadat* het leren zou hebben plaatsgevonden. On-line maten zouden meer kunnen onthullen over het statistischleerproces in verschillende groepen participanten. Bovendien hebben "off-line" maten (zoals de testvragen in de woordsegmentatietaak in hoofdstuk 2) het nadeel dat ze vaak erg moeilijk zijn voor kinderen en geheugen en aandacht waarschijnlijk invloed hebben op hoe goed kinderen deze vragen kunnen beantwoorden. Daarom hebben we on-line maten van statistisch leren toegevoegd aan de woordsegmentatietaak (hoofdstuk 2) en de cross-situational word learning taak (hoofdstuk 3). Bij de woordsegmentatietaak wilden we on-line leren meten met behulp van een klikdetectietaak. Klikgeluiden werden aan de woordenstroom toegevoegd op verschillende plekken: ofwel tussen woorden (*kiba!dalu*), ofwel binnen een woord (*ki!ba*). De proefpersonen moesten zo snel als ze konden op een knop drukken zodra ze een klik hoorden. De hypothese was dat als proefpersonen kennis hadden opgedaan over de woordgrenzen in de stroom, de kliks binnen woorden onverwachter zouden zijn (na *ki* verwachten proefpersonen *ba* en geen klik) dan wanneer een klik tussen twee woorden voorkwam, en dit zou zorgen voor langere reactietijden voor kliks binnen woorden dan voor kliks tussen woorden. Onze verwachtingen kwamen op groepsniveau niet uit. Sommige proefpersonen lieten inderdaad het verwachte effect zien, maar anderen lieten een effect in de tegenovergestelde richting zien. Bij de cross-situational word learning taak hebben we de oogbewegingen van de proefpersonen gemeten tijdens de leerfase, als een on-line maat van statistisch leren. We verwachtten dat als proefpersonen leerden welke woorden en referenten bij elkaar hoorden, ze tijdens de leerfase meer naar het goede plaatje zouden kijken ten opzichte van de afleider. De kinderen keken over het algemeen inderdaad significant meer naar het goede plaatje, maar dit nam niet significant toe tijdens het experiment. Ook vonden we geen significant verschil tussen de kinderen met een TT en de kinderen met een TOS. We kunnen dus niet concluderen dat deze twee groepen kinderen een andere leercurve hebben.

In hoofdstuk 6 worden de resultaten van de verschillende studies besproken. We concluderen dat kinderen met een TOS waarschijnlijk moeite hebben met het gebruiken van statistisch leren om woorden aan referenten te koppelen. Dit zou nadelig kunnen zijn voor de ontwikkeling van lexicaal-semantische kennis – alhoewel we deze hypothese niet kunnen onderschrijven met bewijs voor een relatie tussen impliciet cross-situational word learning en lexicaal-semantische kennis in kinderen met een TOS. We vonden geen bewijs voor een stoornis in visueel distributioneel leren. Dit hoeft niet te betekenen dat de stoornis in statistisch leren in kinderen met een TOS zich beperkt tot statistisch leren waarin taal auditief wordt aangeboden. Om de omvang van de stoornis te onderzoeken, zouden systematische vergelijkingen gemaakt moeten worden van statistischleervermogen in verschillende domeinen, modaliteiten en onderliggende statistische structuren. Onze onderzoeksvraag over de relatie tussen statistisch leren en lexicaal-semantische kennis kan niet beantwoord worden op basis van onze resultaten. Toekomstig onderzoek zou zich kunnen richten op het testen van grotere groepen proefpersonen, het construeren van ecologisch valide statistischleertaken en het vinden van manieren om statistischleervermogen op individueel niveau betrouwbaarder te meten.

About the author

Iris Broedelet was born on January 23rd 1990, in Amsterdam, the Netherlands. She graduated from Fons Vitae Lyceum in 2008, and started the Bachelor Linguistics at the University of Amsterdam in 2009, with a minor in Language Psychology. During her Bachelor she was already greatly interested in the areas of language acquisition, language disorders and psycholinguistics. She graduated cum laude in 2013, and continued with the Research Master Linguistics at the University of Amsterdam in 2014. During this two-year program, she could deepen her knowledge about language acquisition and disorders, and participated in various research projects as well in an clinical internship at Royal Kentalis. She wrote her thesis on procedural learning in adults with dyslexia. After graduating in 2016, she started a position of research assistant at the Linguistics department of the University of Amsterdam, assisting researchers with designing experiments, conducting stimuli and recruiting and testing participants. In 2017 she started her PhD project "Lexical-semantic difficulties in developmental language disorder: the role of implicit statistical learning" under supervision of prof. dr. Judith Rispens and prof. dr. Paul Boersma, after receiving funding from the "PhDs in the Humanities" programme of the Dutch Research Council (NWO). During her project, she presented her research at national and international conferences and published papers in international journals. She has also been a board member of a Dutch association for applied Linguistics (het WAP) since 2018.