Awareness and instruction when kindergarteners acquire grammar

Published by
LOT
Kloveniersburgwal 48
1012 CX Amsterdam
The Netherlands

phone: +31 20 525 2461

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

Cover illustration: Elianne Koolstra.

ISBN: 978-94-6093-396-7
DOI: https://dx.medra.org/10.48273/LOT0612
NUR: 616

Awareness and instruction when kindergarteners acquire grammar

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op 4 februari 2022, te 13.00 uur

door
Sybren Bastiaan Spit
geboren te Valencia

Promotiecommissie

Promotores:	Prof. dr. E.O. Aboh	Universiteit van Amsterdam
	Prof. dr. J.E. Rispens	Universiteit van Amsterdam
Copromotor:	dr. S.J. Andringa	Universiteit van Amsterdam
Overige leden:	Prof. dr. F.P. Weerman	Universiteit van Amsterdam
	Prof. dr. W.B.T. Blom	Universiteit Utrecht
	Prof. dr. E. Marsden	University of York
	Prof. dr. P.J. Monaghan	University of Lancaster
	dr. I. Visser	Universiteit van Amsterdam
	dr. J. Verhagen	Universiteit van Amsterdam

Faculteit der Geesteswetenschappen

Voor Jo

Je gaat het pas zien als je het door hebt

Johan Cruyff

Table of contents

Contributions

Chapter 1 – Introduction

Written by Sybren Spit with valuable feedback from Enoch Aboh, Sible Andringa and Judith Rispens.

Chapter 2 – A miniature language experiment

Chapter 2 is a slightly modified version of a manuscript that is currently under review: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (under review at *Journal of Psycholinguistic Research*). Kindergarteners use cross-situational statistics to infer the meaning of grammatical elements.

The study was designed by Sybren Spit in collaboration with Enoch Aboh, Sible Andringa and Judith Rispens. Dirk Jan Vet assisted with the technical implementation of the experiment. Sybren Spit recruited participants and collected their data. Data analysis was performed by Sybren Spit, primarily supervised by Sible Andringa. Sybren Spit is the lead author of this manuscript, with helpful feedback from Enoch Aboh, Sible Andringa and Judith Rispens.

Chapter 3 – The opt out paradigm

Chapter 3 is a slightly modified version of a published article: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (2019). The opt out paradigm: First steps towards a new experimental method that measures meta-linguistic awareness. *Dutch Journal of Applied Linguistics, 8*(2), 206-227.

The study was designed by Sybren Spit in collaboration with Enoch Aboh, Sible Andringa and Judith Rispens. Dirk Jan Vet assisted with the technical implementation of the experiment. Sybren Spit recruited participants and collected their data. Data analysis was performed by Sybren Spit, primarily supervised by Sible Andringa. Sybren Spit is the lead author of this manuscript, with helpful feedback from Enoch Aboh, Sible Andringa and Judith Rispens.

Chapter 4 – Do kindergarteners develop awareness of statistical regularities?

Chapter 4 is a slightly modified version of a published article: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (2021). Do kindergarteners develop awareness of the statistical regularities they acquire? *Language Learning, 71*(2), 573-611.

The study was designed by Sybren Spit in collaboration with Enoch Aboh, Sible Andringa and Judith Rispens. Dirk Jan Vet assisted with the technical implementation of the experiment. Sybren Spit recruited participants and collected their data. Data analysis was performed by Sybren Spit, primarily supervised by Sible Andringa. Sybren Spit is the lead author of this manuscript, with helpful feedback from Enoch Aboh, Sible Andringa and Judith Rispens.

Chapter 5 – The effect of explicit instruction

Chapter 5 is a slightly modified version of a published article: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (2021). The effect of explicit instruction on implicit and explicit linguistic knowledge in kindergarteners. *Language Learning and Development*, https://doi.org/10.1080/15475441.2021.1941968

The study was designed by Sybren Spit in collaboration with Enoch Aboh, Sible Andringa and Judith Rispens. Dirk Jan Vet assisted with the technical implementation of the experiment. Joris Wolterbeek (student assistant) and Sybren Spit recruited participants. Joris Wolterbeek collected their data. Data analysis was performed by Sybren Spit, primarily supervised by Sible Andringa. Sybren Spit is the lead author of this manuscript, with helpful feedback from Enoch Aboh, Sible Andringa and Judith Rispens.

Chapter 6 – The effects of instruction, input and sleep

Chapter 6 is a slightly modified version of a manuscript that is under review: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (under review at *Language Development Research*). The effects of instruction, input and sleep during kindergartener's acquisition of grammatical structures: an eye tracking study.

The study was designed by Sybren Spit in collaboration with Enoch Aboh, Sible Andringa and Judith Rispens. Talitha Eikenhout (student assistant) and Sybren Spit recruited participants. Talitha Eikenhout collected their data. Data analysis was performed by Sybren Spit, primarily supervised by Sible Andringa. Sybren Spit is the lead author of this manuscript, with helpful feedback from Enoch Aboh, Sible Andringa and Judith Rispens.

Chapter 7 – Discussion

Written by Sybren Spit with valuable feedback from Enoch Aboh, Sible Andringa and Judith Rispens.

Acknowledgements

I would like to thank all participants and participating schools for their help in conducting this research.

Chapter 1

Introduction

Acquiring a language encompasses several seemingly distinct tasks: learners need to detect words within a continuous stream of speech, they need to map meanings onto these words and they have to group these words into abstract categories and determine the grammatical relations between these categories. There are several theories on how learners achieve these tasks. Whereas scholars working from a modular point of view have argued that learning words involves a different cognitive capacity than acquiring grammatical rules (e.g., Berwick & Chomsky, 2015; Chomsky, 1965; Pinker, 1994, 1997), others have suggested that domain general processing capacities, which are solely used for linguistic information, enable learners to grasp the linguistic regularities in all stages of acquisition (e.g., Bybee & McClelland, 2005; Tomasello, 2003; Ullman, 2016). Extending the domain general view, some argue that language acquisition relies on a general cognitive capacity to detect distributional properties within language (e.g., Romberg & Saffran, 2010; Erickson & Thiessen, 2015; Frost & Monaghan, 2016).

This general learning capacity is often referred to as 'statistical learning'. Within statistics, this term refers to a set of mathematical approaches that scientists can use to estimate parameters from data sets (James, Witten, Hastie & Tibshirani, 2013; Hastie, Tibshirani & Friedman, 2001), for example the parameters that can be used to transform a given input X to a particular output Y. Similarly, in cognitive science, statistical learning is regarded as a process of extracting information from distributional properties in the input (Christiansen, 2019). Statistical learning could be seen as a cognitive association mechanism that operates on different stimuli, and enables learners to create mental representations about patterns in the input, which in turn are used to generate relevant behavioral output. Such a statistical learning process has been demonstrated to play an important role in visual learning (e.g., Baker, Olson, Behrmann, 2004; Bertels, Boursain, Destrebecqz & Gaillard, 2015), but has also been suggested to contribute to language acquisition. This contribution has been shown in studies that relate statistical learning capacities to language skills in clinical populations (for developmental language disorders see for example Evans, Saffran & Robe-Torres, 2009 and Lammertink, Boersma, Wijnen & Rispens, 2017 for a meta-analysis; for developmental dyslexia see for example Gabay, Thiessen & Holt, 2015 and van Wittenloostuijn, Boersma, Wijnen & Rispens, 2017 for a meta-analysis), and from computational studies, which show that computational models that learn on the basis of distributional statistics might reflect the same language learning process that humans go through (Aslin & Newport, 2014; Mintz, Newport & Bever, 2002; Qian, Reeder, Aslin, Tenenbaum & Newport, 2012; Swingley, 2005).

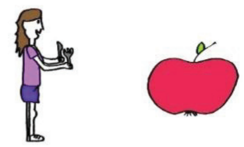
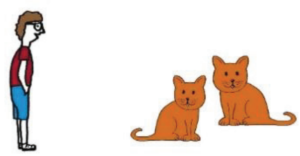
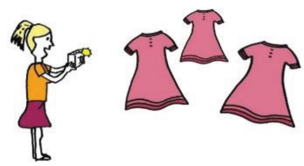
Much insight into statistical learning comes from experiments in which learners need to use different types of statistical information to detect relationships between linguistic elements. So-called artificial grammar learning experiments have shown that tracking statistical regularities allows learners to infer word boundaries and establish dependencies between linguistic elements (e.g., Aslin, Saffran & Newport, 1998; Gómez & Gerken, 1999; Saffran, Johnson, & Aslin, 1996 for infants; Gómez, 2002 for infants and adults; Endress & Bonatti, 2007; Perruchet & Vinter, 1998; Thiessen, Kronstein, & Hufnagle, 2013 for adults). Cross-situational word-referent learning tasks indicate that statistical learning plays a role in mapping form onto meaning (e.g., Frank, Goodman & Tenenbaum, 2009; Frank, Tenenbaum & Fernald, 2013; Smith, Suanda & Yu, 2014; Smith & Yu, 2008; Vlach & Johnson, 2013 for infants; Kachergis, Yu, & Shiffrin, 2014; Vouloumanos, 2008; Yu & Smith, 2007 for adults). Several studies also show that distributional properties in the input can help learners to determine the semantic category words belong to (e.g., Lany, 2014; Lany & Saffran, 2013 for infants; Chen, Gershkoff-Stowe, Wu, Cheung & Yu, 2017; Monaghan, Mattock, Davies & Smith, 2015 for adults).

In short, statistical learning seemingly explains many aspects of language acquisition, but there are several open questions, some of which we aim to contribute to in this dissertation. In the remainder of this introduction, we will introduce some of these questions by describing the statistical learning experiment that forms the foundation of the studies in this book. With this description, we aim to identify gaps in our understanding about statistical learning and its contribution to young children's language acquisition process.

1.1 A statistical learning experiment

In our experiments, kindergarteners had to acquire a meaningful grammatical pattern from a miniature language. Children were told a story about four protagonists (Carlo, Julia, Marco and Maria), who were going on a holiday to a country where they did not speak the language. Participants were asked to help the protagonists learn this new language. They were told they would see pictures and hear things in the new language that matched the pictures they saw. Sentences in the language had subject-verb-object word order, as Dutch has in main clauses. In these sentences, nouns must be introduced by a nominal marker, unlike Dutch. The language had two of these markers: *pli* and *tra*. A noun introduced by *tra* can refer to both singular and plural referents. The correct interpretation must be inferred from the visual context. *Pli*, however, encodes number and indicates that the noun necessarily refers to multiple referents. In short, the rule participants had to learn was that whenever *pli* preceded a noun, this noun always referred to multiple referents. See Table 1.1 for an example of these different types of sentences.

Table 1.1. Examples of artificial language training sentences and rough translations.

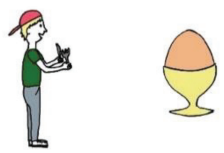
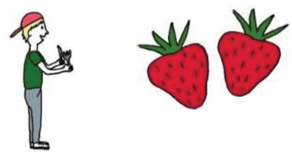
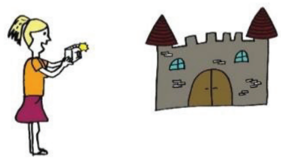
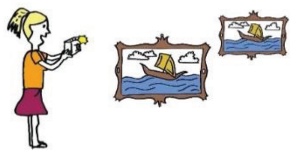
	Example	Translation	Picture
***Tra* + single referent**	*Maria rigarda tra zambo*	'Maria eats (an) apple(s)'	
***Tra* + multiple referents**	*Carlo estima tra pano*	'Carlo looks at (a) cat(s)'	
***Pli* + multiple referents**	*Julia pentura pli anso*	'Julia takes a picture of dresses'	

Note: For sentences with *tra* the noun could be translated both as a singular and a plural. The correct interpretation should be inferred from the visual context.

The experiment consisted of three parts. It started out with a short vocabulary training. After this, a rule training session followed. During this session, children received input to acquire the grammatical rule by hearing sentences like those in Table 1.1. The statistical regularity in the input was the following: whenever children hear *pli,* chances of seeing multiple objects were a hundred percent, but whenever they heard *tra,* this chance was fifty percent. Children thus had to learn that the grammatical marker *pli* indicated plurality. After the rule training, participants performed a test phase to test whether they acquired this regularity. The test phase was followed by an exit interview.

Children were tested by means of a picture matching task to determine whether they became sensitive to the grammatical structure. In this task, participants heard sentences from the rule training phase and had to choose which of two pictures matched each sentence. In the experimental test items of this task, participants had to choose between pictures with either one or multiple referents. The two pictures always referred to different referents. Half of the experimental items contained *pli*. For these items, the target picture always showed multiple referents. The other half of the experimental items contained *tra*. Any number of referents would be grammatical for these items. Sensitivity to the statistical regularity in the input would lead to more target answers on trials with *pli* than on trials with *tra*, because for trials with *pli* participants could base their responses on the number of referents the picture showed, whereas the number of referents was not indicative for trials with *tra*. Examples of the different types of test items can be seen in Table 1.2.

Table 1.2. Examples of test items and their rough translations.

	Marco rigarda pli bovo - 'Marco eats strawberries'	
Pli		
	Julia pentura tra nutro – 'Julia takes a picture of (a) painting(s)	
Tra		

Note: For sentences with *tra* the noun could be translated both as a singular and a plural. In both examples, the right picture was the target.

In the studies reported in this thesis, which made use of this setup, children generally gave more correct answers when the trial included *pli* than when it included *tra*. This might be taken as evidence that young children have a discovery procedure that makes them sensitive to such grammatical regularities, as the distributional properties in the input were the only cue to detect this regularity. Possibly, children employ statistical learning mechanisms to acquire a grammatical marker that also carries meaning. As mentioned before, studies have already shown that statistical learning plays a role in word segmentation (e.g., Aslin, et al., 1998; Endress & Bonatti, 2007; Gómez, 2002; Gómez & Gerken, 1999; Perruchet & Vinter, 1998; Saffran, et al., 1996; Thiessen, Kronstein, &

Hufnagle, 2013), word referent learning (e.g., Frank, et al., 2009; Frank, et al., 2013; Kachergis, et al., 2014; Smith, et al., 2014; Smith & Yu, 2008; Vlach & Johnson, 2013; Vouloumanos, 2008; Yu & Smith, 2007) and the acquisition of semantic categories (e.g., Chen, et al., 2017; Lany, 2014; Lany & Saffran, 2013; Monaghan, et al., 2015). Although the statistical information that learners needed to use varied considerably between these studies, learners always had to associate particular pieces of information with each other. What has not been studied, is whether this is the same with grammatical markers that express meaning, such as markers of nominal number. In the study that will be discussed in more detail in Chapter 2 we used the setup we just described to test whether kindergarteners could grasp a statistical dependency between an auditory linguistic element and a visual plural interpretation of this marker. Results from this study are compatible with the idea that statistical learning plays a role in learning such meaningful grammatical elements. In Chapter 2, we will also elaborate more on why this structure is of interest for studies on statistical learning. Chapter 2 thus focusses on the first main question of this book: is statistical learning involved in acquiring grammatical structures that carry meaning?

1.2 How implicit is statistical learning?

Interestingly, when children participated in an exit interview after the test phase of the miniature language learning experiment, they could not verbalize any knowledge of the grammatical marker they just learned. Not a single child could say that *pli* was used to indicate plurality. Results like this are often taken as an indication that statistical learning is an implicit learning process (e.g., Reber, 1967; Aslin, et al., 1998; Arciuli & Simpson, 2012; Kidd, 2011), because children seemingly lack awareness of the grammatical structure they have acquired. Implicit learning occurs when learners are not aware that they are learning and do not seem to make a conscious effort to discover the regularities in the target language, whereas explicit learning happens when learners know they are learning and do have such intentions (e.g., Dörnyei, 2009; N. Ellis, 2015; Hulstijn, 2015; Ullman 2001, 2016). Often, these types of learning are defined by the type of knowledge they result in (e.g., Hulstijn, 2015; Reber 1967), with explicit learning seen as learning that results in explicit knowledge and implicit learning as learning that results in implicit knowledge. Explicit knowledge can be described as knowledge that is available to awareness, whereas implicit knowledge is not available to our awareness (Rebuschat, 2013).

Rebuschat and Williams (2012) observed that statistical learning and implicit learning are sometimes equated (Conway & Christiansen, 2006; Perruchet & Pacton, 2006; Turk-Browne, Jungé & Scholl, 2005) and that from time to time authors concatenate the two when they speak of 'implicit statistical learning' (e.g., Conway, Bauernschmidt, Huang & Pisoni, 2010; Kidd, 2011). However, statistical learning does not need to be implicit *per se* (Batterink, Reber, Neville & Paller, 2015; Bertels et al., 2015), and could happen both implicitly and explicitly. As statistical learning likely plays an important role in child language acquisition (e.g., Erickson & Thiessen, 2015; Frost & Monaghan, 2016; Romberg & Saffran, 2010), it is often assumed that children acquire a target language without any awareness of the regularities in it. This might be a reasonable assumption, because children are probably unable to verbalize or explicitly reflect on their knowledge, which may be why we found no evidence of this in the exit interviews, as the results in Chapter 2, and other chapters of this book show. Many scholars indeed make this assumption, regardless of the theoretical framework they adopt (Aslin et al., 1998; Chomsky, 1986; Radford, 2004; Saffran et al., 1996; Tomassello, 2003; Ullman, 2016; Wijnen, 2013).

It remains to be seen, however, whether this lack of verbalization necessarily reflects children's unawareness of the acquired knowledge. Instead, one of the ideas put forward in this book is that children might be aware of the knowledge they acquire even though they cannot describe this knowledge in words, and that research so far has failed to tap into awareness at this level. Learners may possess phenomenal awareness (P-awareness), when they can verbalize their experiences, or could be aware at the level of access (A-awareness), when they cannot verbalize or remember their experiences coherently (Cleeremans, 2008, 2011, 2014). A verbal report as used in this experiment only taps into the former type of awareness (Timmermans & Cleermans, 2015), and is thus inconclusive about the involvement of awareness of the latter type. Perhaps, participants in the experiment that is described in Chapter 2 were aware of the regularity in a way they could not verbalize.

Research with adults shows that measuring awareness is a difficult enterprise, and several measures of awareness can be distinguished (see Rebuschat, 2013 for a review). Traditional methods of investigating meta-linguistic awareness, such as verbal reports or confidence judgement tasks, typically measure P-awareness (Timmermans & Cleeremans, 2015; Batterink, et al., 2015) and thus seem unsuitable to investigate awareness in children. Therefore, we adopted the 'opt out paradigm', which has been used to investigate (A-)awareness in animals (Hampton, 2001, 2009). The opt out paradigm assesses awareness without requiring learners to verbalize their experiences. In this paradigm participants show strategic behavior that reveals their awareness. We believe this method can also be used to assess whether children are aware of the linguistic patterns they are acquiring.

In Chapter 3, we explore the possibility of using the opt out paradigm as a measure of meta-linguistic awareness in children. The chapter also provides a more elaborate theoretical account of how we could conceptualize awareness. The results from this chapter suggest that the opt-out method can indeed be used to gain more insight into whether young learners, who are unable to verbalize awareness of acquired linguistic structure, could still possess such awareness. In this regard, Chapter 3 presents an exploration of this potentially suitable method.

No conclusive results emerge, however, about whether children indeed do develop such knowledge. To answer this question, we appended the opt out experiment to the miniature language experiment that we already briefly discussed in Section 1.1, and which is presented in more detail in Chapter 2. Chapter 4, thus, combines the artificial language learning task presented in Chapter 2, and the opt out procedure from Chapter 3. Results presented in Chapter 4 suggest that children may indeed develop A-awareness when they acquire a grammatical pattern on the basis of statistical regularities in the input. Even though young children cannot verbalize their awareness of acquired linguistic knowledge, they nevertheless may be aware of this knowledge.

1.3 Does awareness matter?

One may wonder why it matters whether children develop and possess awareness of learned linguistic patterns or not. Within linguistics, awareness has been studied most extensively within second language acquisition (SLA), often focusing on the earlier mentioned distinction between implicit and explicit knowledge (e.g. Andringa & Rebuschat, 2015; Hulstijn, 2015; Suzuki & DeKeyser, 2017). The research and findings on this topic are diverse (e.g. DeKeyser, 2003; Ellis, 2015; Han & Finneran, 2014; Paradis, 2009), and have led to the formulation of several theories about the role that awareness plays during the acquisition of the grammatical structures of a new language. Some authors claim that awareness of such structures in a second language is necessary to acquire them (DeKeyser, 2003; Schmidt, 1990), whereas others suggest awareness merely facilitates the acquisition process (Ellis, 2003). Likewise, other researchers have argued awareness is only circumstantial and a consequence of how the language is learned, but has no influence on the implicit linguistic knowledge that is gained (Krashen, 1981).

The role of awareness in early language acquisition has been studied too, but perhaps not as extensively as in SLA. Many studies have investigated the role of awareness and explicit knowledge in the acquisition of vocabulary knowledge (e.g. Coyne, McCoach, Loftus, Zipoli & Kapp, 2009; Silverman, 2007; Vaahtoranta, Suggate, Jachmann, Lenhart & Lenhard, 2018). A meta-analysis showed that vocabulary interventions in kindergarteners are effective and that more explicit instruction conditions result in bigger effect sizes (Marulis & Neuman, 2010). Furthermore, phonological awareness is widely studied in young children (e.g. Degé & Schwarzer, 2011; Gillon, 2018), and is often stressed as an important contributor to reading (e.g. Furnes & Samuelsson, 2011; Johnson & Goswami, 2010; Saygin et al. 2013). Nevertheless, the general theoretical debate about the role of awareness during language learning that is held in SLA seems absent in the field of child language acquisition in general, and statistical learning in particular.

The absence of this debate appears to lie in the fact that some scholars argue that children learn grammatical rules using implicit learning mechanisms only (Bialystok, 1994; DeKeyser & Larson-Hall, 2005; Paradis, 2004, 2009). This is perhaps best illustrated by Bialystok (1994) who posits "Rules make sense to adults; they make little difference to young children" (p.565), suggesting children do not resort to explicit learning when acquiring grammatical structures. Lichtman (2016) captures this idea in the 'maturational hypothesis' which posits that both children and adults are maturationally constrained to using particular learning mechanisms. This is a take on the idea that there is a fundamental difference between young children who learn the grammatical regularities of a new language and adults who go through the same process (Bley-Vroman, 1990; Hartshorne, Tenenbaum & Pinker, 2018; Krashen, Scarcella & Long, 1982; Singleton & Ryan, 2004): there would be a difference in the available learning mechanisms for adult and child language learners (Lichtman 2016; Pakulak & Neville, 2011; Zwart, Vissers, Kessels & Maes, 2019). In this scenario, children use more implicit or procedural learning mechanisms when they acquire grammatical structures, whereas adults rely on explicit or declarative learning mechanisms, and possibly in conjunction with implicit learning mechanisms (DeKeyser, 2000, 2003; R. Ellis, 2005, 2009; Paradis, 2004, 2009; Ullman, 2001).

However, hardly any studies have investigated specifically whether young children actually can learn grammatical patterns using explicit learning mechanisms. It could well be the case that children resort only to implicit learning, but to our knowledge such claims have rarely been tested experimentally. In Chapters 5 and 6, we begin to fill this void by investigating if young children benefit from explicit instruction of a grammatical regularity. We conducted the same artificial language learning experiment as in Chapters 2 and 4, but to test whether young children can make use of explicit learning mechanisms, we investigated whether explicit instruction influences the acquisition of the grammatical marker. In Chapters 2 and 4, we probed into children's knowledge using a picture matching task only, which is considered a more explicit measure in second language acquisition. The experiments in Chapters 5 and 6 also followed children's eye movements during the task to gauge knowledge using a more implicit measure as well. Measures like picture matching tasks are often more associated with explicit knowledge, whereas eye tracking is assumed to tap more into implicit knowledge (Andringa & Rebuschat, 2015; Bowles, 2011; R. Ellis, 2005; Godfroid, Loewen, Jung, Park, Gass & Ellis, 2015; Han & Ellis, 1998). As explicit instruction could affect the development of these types of knowledge in different ways (Godfroid, 2015), it was important to include these two different measures (e.g., Andringa, de Glopper & Hacquebord, 2011).

In the experiment in Chapter 5, children were exposed to the miniature language again. Importantly, half of the children received only this exposure, whereas the other half received explicit instruction about the marker in addition to the exposure. Afterwards, all children were tested on their knowledge using a picture matching task during which we measured their eye movements. Chapter 6 presents a replication of the study from Chapter 5, but where children were tested twice: once directly after the exposure, and a second time on the day after their first test. This way we could get a clearer picture of how explicit instruction affects acquisition of a grammatical marker and how this process is influenced by (a) a consolidation period including sleep, and (b) by receiving more input to learn the grammatical marker. Consolidation is defined as the process during which information that is stored in our memory gets strengthened and enhanced (Axelsson, Williams & Horst, 2016; Dewar, Alber, Cowan & Della Sala, 2014; Dudai, 2002). Consolidation can occur when a learner receives more input, and encoded information becomes represented more strongly (Ellis, 2002; Bybee, 2010), but wakeful rest and sleeping time after encountering a stimulus can also be beneficial for enhancing stored information (Dewar, Alber, Butler, Cowan & Della Sala, 2012; Mednick, Cai, Shuman, Anagnostaras & Wixted, 2011; Wixted, 2004 for wakeful rest; Diekelmann, Wilhelm & Born, 2009; Stickgold & Walker, 2013 for sleep).

Results from Chapters 5 and 6 did not provide any evidence that explicit instruction leads to higher accuracy rates, but suggested there was evidence of changed eye movement behavior because of such instruction. Conversely, receiving more input seemed to have an effect on accuracy scores, but did not influence eye movements. This might mean that young learners can use explicit instruction to a certain extent, but that it leads to rather unstable representations, whereas increased input leads to more robust knowledge in this age group. Explicit instruction appears to affect learning in young children, but potentially not in the same way as is typically observed in adults. If explicit instruction plays a role in child language acquisition, a possible effect of explicit instruction is most likely modulated by many other factors, like the linguistic background or general cognitive capacities of these children. Future research needs to address how such factors interact with explicit instruction in these young learners, to get a better idea what the role of explicit instruction in this age group precisely is.

1.4 Overview

What we aim to show with the studies in this book is that the role awareness plays in early language acquisition is a highly relevant, yet under-investigated topic. The conclusion of this book is that although awareness probably does not play the same role in the learning process of children as it does in adults, this does not necessarily mean it plays no role whatsoever. As such, it is a topic worth investigating more deeply. However, before we reach that conclusion the interested reader might want to learn more about the experiments and data that try to justify this claim. Chapter 2 presents the miniature language experiment that most studies in this book have been built onto. Ample details about the method will discussed in this chapter in order to familiarize the reader with the discussion in subsequent chapters. Chapter 3 discusses awareness in greater detail and presents a novel method to measure awareness in young children. Chapter 4 uses the novel method and combines it with the miniature language from Chapter 2 to investigate whether young learners develop awareness of the grammatical structures they acquire. Chapters 5 and 6 then further investigate whether kindergarteners also use awareness of grammatical structures when acquiring them, by investigating the effect of explicit instruction on learning in this age group, again using the same method from Chapter 2 as a basis. Chapter 7 expands on some of the bigger questions that have been raised in this introduction, and also presents some concluding remarks, based on the findings in preceding chapters.

One final point that might have become clear from this overview is that the same miniature language experiment is used in several chapters. This is not without a reason. In the spirit of recent calls for more replication (Marsden, Morgan-Short, Thompson & Abugaber, 2018), a secondary goal of this project was to determine the robustness of the statistical learning effect that we were after. We therefore repeatedly used the same method, and encourage others to do the same. All materials, and scripts used for analysis can be consulted online, on our OSF page (https://osf.io/bp5qe/; Spit, Andringa, Rispens, Aboh & Vet, 2019). Hopefully they can be used by others either for inspiration, replication, or to develop further and improve.

Chapter 2

A miniature language experiment[*]

In the introduction of this book, we briefly discussed how several studies demonstrate that detecting statistical regularities in the linguistic input plays a key role in language acquisition. The process of extracting information from distributional properties in the input is often labeled statistical learning. Statistical learning has been demonstrated to play an important role in visual learning (e.g., Baker et al., 2004; Bertels et al., 2015), but it has also been suggested to contribute to language acquisition. This contribution has been shown in studies that relate statistical learning capacities to language skills in clinical populations (for developmental language disorders see for example Evans, et al., 2009 and Lammertink et al., 2017; for developmental dyslexia see for example Gabay et al., 2015 and van Wittenloostuijn et al., 2017). Further evidence for a relationship between statistical learning and natural language comes from computational studies. Such studies, which often combine computational modeling with corpus research, show that computational models that learn on the basis of distributional statistics reflect natural language learning processes (Aslin & Newport, 2014; Mintz et al., 2002; Qian et al., 2012; Swingley, 2005). Although there are good reasons to assume that statistical learning plays an important role in language acquisition, it has been argued that statistical learning capacities cannot solely explain how languages are acquired (Lidz & Gagliardi, 2015; Yang & Montrul, 2017). The question is rather to what extent statistical learning is involved in the

---

[*] This chapter is a slightly modified version of a manuscript that is currently under review: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (under review at *Journal of Psycholinguistic Research*). Kindergarteners use cross-situational statistics to infer the meaning of grammatical elements.

acquisition process. Later chapters in this book are devoted to factors that might interact with statistical learning, such as awareness and consolidation. This chapter presents an experimental method by which we investigated whether statistical learning is involved in acquiring an agreement marker that carries additional meaning (i.e., beyond the meaning expressed by the word it agrees with), which is a property of many natural languages. We further build on this same method in the experiments in Chapters 4, 5 and 6.

2.1 Statistical learning experiments

Studies within the artificial grammar learning paradigm have shown that tracking statistical regularities allows learners to infer word boundaries and establish dependencies between linguistic elements (e.g., Saffran et al., 1996; Endress & Bonatti, 2007; Gómez & Gerken, 1999; Perruchet & Vinter, 1998, Thiessen et al., 2013). In such studies, participants are typically confronted with a carefully designed stream of artificial linguistic input, with certain elements co-occurring and others not. After exposure, participants are tested on their ability to distinguish between stimuli that co-occurred in the input and stimuli that did not co-occur systematically. If participants were able to track the statistical dependencies between co-occurring elements during exposure, they should be able to recognize those stimuli during such a test. Previous results suggest infants detect such regularities (Aslin et al., 1998), and it has also been shown that they are able to generalize acquired regularities to novel items, when co-occurring elements are not directly adjacent (Gómez, 2002). It is important to note that, unlike in natural languages where auditory input possesses meaning, the statistical regularities in these studies did not carry meaning.

Recent evidence seems to indicate that statistical learning also plays a role in form-reference mappings (e.g. Frank et al., 2013; Frank et al., 2009;

Kachergis et al., 2014; Smith et al., 2014; Vouloumanos, 2008; Yu & Smith, 2007). This has been established using cross-situational statistical word referent learning tasks, in which participants have to acquire nonsense words and their meanings. Instead of tracking which linguistic sounds co-occur together, learners now have to grasp the statistical dependency between certain sounds and particular visual stimuli. In a typical trial, participants are exposed to multiple pictures (e.g. a dog and a cat) simultaneously and hear the accompanying nonsense words (e.g., *bovo* and *pano*) in a sound stream. In a subsequent trial, they are presented with a novel combination of pictures (e.g. a horse and a cat) and matching words (e.g., *pano* and *orbo*). Although it is impossible to infer which word refers to which picture from a single trial, this can be accomplished across multiple trials if a participant is able to detect that a certain nonsense word (*pano*) always occurs in combination with a particular picture (a cat). Studies report that infants are able to infer word meanings in such tasks (e.g. Smith & Yu, 2008) and that, as they grow older, they are capable of inferring word meanings when the interval between word referent pairs increases (Vlach & Johnson, 2013).

Furthermore, studies like those by Lany (2014) and Lany and Saffran (2013) indicate that distributional properties in the input help learners to determine the semantic category words belong to. They showed this using cross-situational learning tasks in which referents were divided into two categories (e.g., animals and vehicles) and words that referred to them were accompanied by different suffixes/particles (e.g., *erd* and *alt*) or had different syllabic structure (e.g., mono- and bi-syllabic). At test, participants were exposed to a new word that had the same characteristics as words they encountered during training (*erd* and mono-syllabic) and were asked to match this new word to a picture of a particular category (animals). Participants succeeded to link the right properties to the right category (i.e., that mono-syllabic words refer to animals). In these experiments, participants would thus show learning when they were able to detect the statistical association between linguistic elements and corresponding abstract semantic categories. Similar experiments have also shown that learners can use such information to categorize nouns and verbs (Monaghan et al., 2015), to group objects according to their shape (Chen et al., 2017), and to determine which words in the input are function words and which are not (Hochmann, Endress & Mehler, 2010).

Studies in the lab have further shown that learners can acquire more complex grammatical patterns aided by the distributional properties of the input (e.g., Goldberg, Casenheiser & Sethuraman, 2004; Wonnacott, Brown & Nation, 2017). In these studies, participants were exposed to English sentences containing a novel grammatical structure that they could map to a novel meaning. For example, a sentence with a novel verb and two English nouns in a subject-object-verb order would mean that the subject appeared on or into the object in the manner denoted by the verb (Goldberg, Casenhiser & White, 2007). The statistical regularity participants had to grasp here was the association between a particular word order and a specific verbal meaning. At test, participants heard the grammatical construction with another novel verb, and saw two pictures of which one contained the intended meaning. Participants showed learning by mapping this word order to the correct picture. Similar studies show that 5- to 7-year old children learn the grammatical form better when a single verb occurs more frequently in the new construction than others (Casenheiser & Goldberg, 2005), and that in 6-year-olds the knowledge of phrasal constructions is stronger when they are retested a couple of days later (Wonnacott, Boyd, Thomson & Goldberg, 2012). Importantly, studies of this kind are typically not classified as 'statistical learning studies'. Although the properties of the input are manipulated in these experiments, they involved incorporating a novel structure into the native language of participants, instead of exposing participants to a language in which all lexical items are completely new to them. The exact statistics of the input are therefore not controlled as meticulously as in statistical learning studies.

2.2 The current study

The statistical learning studies discussed above often vary considerably in terms of the statistical information that learners need to draw on: from tracking co-occurrence between sounds (e.g. Saffran et al., 1996; Endress & Bonatti, 2007; Gómez & Gerken, 1999) to matching a word order pattern to a verbal meaning (e.g. Goldberg et al, 2004; Wonnacott et al., 2017). However, they all share some important commonalities. In order to acquire the investigated linguistic structures, learners needed to associate particular stimuli with each other. Whether these stimuli involved syllables, word orders, or abstract categories, participants acquired the linguistic structures by tracking the statistical dependency between these stimuli. It seems reasonable to assume that one domain general capacity for statistical learning is involved in these different learning tasks (Romberg & Saffran, 2010). Statistical learning has thus been demonstrated to be involved in tracking word boundaries, learning words and their referents, but also in grouping words in larger semantic categories and in acquiring more complex grammatical structures.

Yet, languages do not consist of only lexical categories but also involve functional categories (e.g., demonstratives, adpositions, tense, mood, aspect markers) which convey specific grammatical meaning and therefore determine how the lexical category they combine with must be interpreted. The following examples (1) and (2) illustrate the contribution of such grammatical elements in English and Gungbe. In English, a noun that refers to a countable entity cannot occur in a sentence by itself, but needs to be combined with a determiner. This can be a definite article, as in (1a), which enforces the interpretation of 'table' as known to both speaker and hearer, or a demonstrative as in (1b), which denotes proximity to the speaker.

(1a) ‘Koku bought *the* table’

(1b) ‘Koku bought *this* table’

In Gungbe, a Gbe language of the Kwa family, which is spoken in Porto-Novo (Republic of Benin) and in environs of Gbadagri (Republic of Nigeria), we encounter a different situation. This language allows bare nouns to occur in a sentence, with an interpretation that could be (in)definite singular, generic, or generic plural (i.e., a/the table, (some) tables, or tables in general). Only the context allows speakers to tease these different interpretations apart (Aboh 2004, Aboh & DeGraff, 2014). In this language a bare noun (2a) must be distinguished from a noun accompanied by the grammatical element *lɛ́* (2b), which forces the interpretation of the noun as plural and definite.

(2a)	*Kɔ̀kú*	*xɔ̀*	*távò.*	
	Koku	buy	table	
	‘Koku bought a/the table.’			
	‘Koku bought (some) tables.’			
(2b)	*Kɔ̀kú*	*xɔ̀*	*távò*	*lɛ́.*
	Koku	buy	table	NUMB.DEF
	‘Koku bought the tables.’			

When comparing these English and Gungbe examples, we see that the grammatical elements not only determine the grammatical construction of a phrase but also contribute to the meaning of that phrase. Indeed, proximity, definiteness, and number are not inherent semantic properties of the nouns *távò* and *table*. These meanings are added through the grammatical elements the nouns co-occur with, and learners need to grasp this association to interpret these phrases correctly. An important point here is that while lexical items can refer to entities or events that have some representation in the actual world and in such cases allow acquisition of their semantics, the semantics of grammatical items can only be acquired through linguistic context. We therefore wanted to know whether statistical learning is involved in acquiring grammatical or functional elements using a miniature language learning experiment.

This study is not the first to investigate whether young children acquire this type of grammatical structure in an artificial language learning setting. Culbertson and Newport (2015), for example, investigated whether children prefer a particular word order when learning a grammatical marker. They showed children pictures of objects, and combinations of a modifying marker and a noun describing the picture. These combinations could occur in two different orders (e.g. noun-marker and marker-noun), one of which was more frequent in the input. Input frequency did not matter for learning the meaning of the modifier, but was sometimes in line with what would be predicted from linguistic universals, and sometimes not. The researchers found that when children saw a picture and had to describe it, they overgeneralized the more frequent word order for their descriptions only when the order was in line with what linguistic universals would predict. Using a comparable kind of experimental set-up, Tal and Arnon (2020) showed that children are able to learn an optional plural marker, but that children are also less likely to use this optional marking for noun classes that occurred as plurals in the input infrequently than for noun classes that occurred as plurals frequently. Furthermore, Raviv and Arnon (2018) found that in an iterated learning experiment, children could learn a similar grammatical element expressing plurality from exposure to noun phrases, even when linguistic input was relatively unstructured, because earlier generations of learners produced unstructured output.

Although these studies show that young children are able to learn meaningful grammatical elements from limited linguistic input in various artificial language learning settings, Culbertson and Schuler (2019) pointed out that the languages used in many of these studies are rather different from what natural languages look like. Studies like those described above, for example, present children with merely noun phrases (e.g. Culbertson & Newport, 2015; Raviv & Arnon, 2018; Tal & Arnon, 2020). To see whether learners also rely on statistical learning to acquire grammatical elements in more natural settings, it is important to test them using an experimental procedure where the input resembles a natural language more closely than in typical artificial language learning experiments.

In the present experiment, kindergarteners were trained on an artificial language in which a grammatical marker combines with nouns to express number. Our aim was to investigate whether kindergarteners use statistical information from the input to acquire a meaningful grammatical element. In our language, nouns could be preceded by two markers: *pli* and *tra*. The marker *tra* was a default nominal marker with no number specification, whereas *pli* indicated that the noun always refers to multiple referents. Statistical properties of the input were the only cue for acquiring this meaningful grammatical element. To learn this regularity, children were exposed to a miniature language containing novel lexical items of which the statistics were carefully manipulated in order to create a closer resemblance to natural languages than studies we described earlier. The language not only contained noun phrases to acquire the grammatical marker, but also verbs, proper names and conjunctions. At test, children were presented with a picture task during which number was the crucial feature to distinguish between the possible answers. Because the grammatical marker *tra* was not indicative of number, whereas *pli* was, we expected learning the grammatical marker would lead to better performance for test items including *pli* than for test items including *tra*. Thus, our experiment was designed in such a way that we could answer the question whether distributional properties of the input allow children to acquire a grammatical element that expresses number in a language that resembles a natural language more closely than previous artificial grammar learning experiments.

2.3 Method

2.3.1 Participants

50 native Dutch speaking children (25 males, 25 females, *M* = 5;5 years, *SD* = 0;10, *range* = [4;2 - 7;1]) took part in this experiment. All children were in kindergarten and were recruited from a primary school in Haarlem. Their teachers reported that none of the children were diagnosed with any language or communication disorders. No restrictions were imposed upon taking part in the experiment. No further information about the children's (non-)linguistic background was collected.

2.3.2 Materials

2.3.2.1 Miniature language and target structure

To place the children in a naturalistic learning environment, an artificial language was created that consisted of four proper names, three verbs, two grammatical markers, one conjunction, six frequent nouns, and twelve infrequent nouns, which were necessary for designing test items (see below). All words and their translations can be found in Table 2.1. Apart from the proper names, which might occur in Dutch, all words were novel words, which were loosely based on word forms from Esperanto. The chosen forms were sometimes slightly manipulated such that all word forms within a particular word class were equally likely to be Dutch words, based on the transitional probabilities between phonemes within each word.

Table 2.1. All words from the miniature language and their translations.

Word type	Word	Translation
Proper name	*Carlo*	'Carlo'
	Julia	'Julia'
	Marco	'Marco'
	Maria	'Maria'
Verbs	*Estima*	'Looking'
	Pentura	'Taking a picture'
	Rigarda	'Eating'
Frequent nouns	*Domo*	'Tree'
	Herbo	'Banana'
	Kego	'Horse'
	Lito	'Flower'
	Pano	'Cat'
	Zambo	'Apple'
Infrequent nouns	*Ando*	'Carrot'
	Anso	'Dress'
	Arbo	'Castle'
	Bovo	'Strawberry'
	Halto	'Rabbit'
	Kobro	'Sheep'
	Misto	'Dog'
	Nego	'Egg'
	Nutro	'Painting'
	Teko	'Sandwich'
	Wiro	'Car'
	Wolgo	'Cow'
Grammatical marker	*Pli*	'Plural'
	Tra	'Any number'
Conjunction	*Ut*	'And'

The language had subject-verb-object word order, like Dutch has in main clauses. In the miniature language, a noun phrase on its own does not encode number and could correspond to both singular and plural referents. In a sentence, however, an argument noun phrase must be introduced by the nominal marker *tra*. The type of nominal marker included in our artificial language is quite common in languages with residual noun classes, such as the Kwa-language Akan (Appah, 2003) or the Austronesian language Cebuano (Parnes, 2011). In our miniature language, a noun introduced by *tra* can refer to both singular and plural referents: the correct interpretation must be inferred from the visual context. The sole function of *tra* therefore is to turn a bare noun into an argument. The language includes another nominal grammatical marker *pli*, which encodes number and indicates that the noun necessarily refers to multiple referents. Apart from carrying information about definiteness, the function of this marker is similar to the previously discussed marker from Gungbe. The marker *tra* can be seen as a default nominal marker with no number specification, whereas *pli* not only turns the noun into an argument but adds information about number. In short, the rule participants had to learn was that whenever *pli* preceded a noun, this noun always referred to multiple referents. Dutch does not contain such a number marking grammatical element. Instead, Dutch nouns are pluralized through suffixation of the noun (Booij & van Santen, 2017), which is comparable to English pluralization. Accordingly, we did not expect the children to draw on their L1 to acquire the grammatical rules of the miniature language.

2.3.2.2 Structure of the experiment

Before describing the precise characteristics of the input to learn this grammatical regularity in the next section, we will briefly lay out the overall structure of the experiment. An overview of this structure can be seen in Figure 2.1. To learn the regularity, participants were told a story with four protagonists (Carlo, Julia, Marco and Maria), who were going on a holiday to a country of which they did not speak the language. Participants were asked to help the protagonists learn this new language. They were told they would see pictures and hear things in the new language that matched the pictures they saw. The experiment consisted of three parts. It started with a short vocabulary training, followed by a rule training session. During the latter session, children received input to acquire the intended grammatical rule. After the rule training, participants performed a test phase to test whether they acquired the rule.

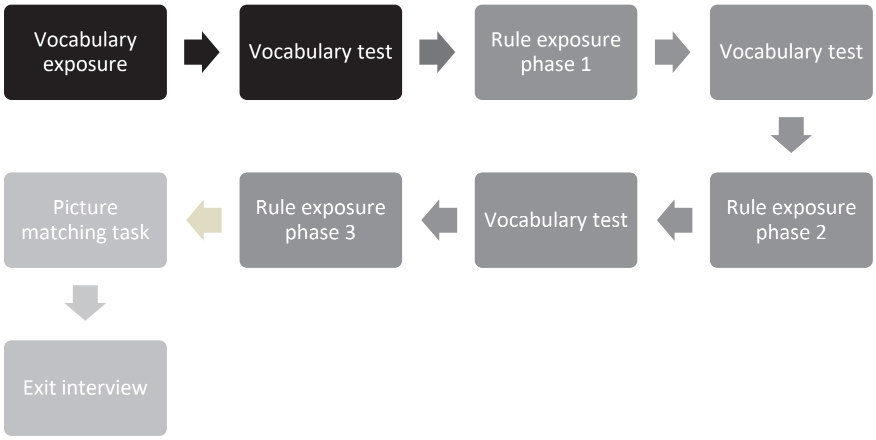
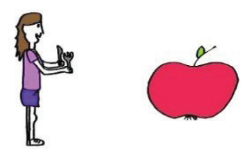
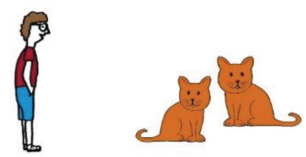
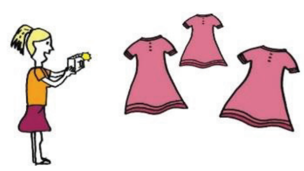


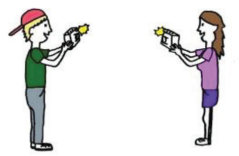
Figure 2.1. A visual representation of the structure of the experiment. The darkest boxes show the parts of the vocabulary training. The lighter boxes show the different parts of the rule training. The lightest boxes show the test phase.

As we were interested in the acquisition of the grammatical rule and function, and not in that of lexical vocabulary items, we bootstrapped participants into learning the lexical items by means of a short vocabulary training featuring just the six frequent nouns of the language. In this training, the nouns were presented auditorily, without the grammatical markers *pli* or *tra* accompanying them, and with a picture showing the meaning of the corresponding bare noun. Each noun was presented six times during this training, twice with one, twice with two and twice with three referents, to make sure a bare noun would not be associated with a particular number of referents. After this training, participants were given a picture matching task to test their vocabulary knowledge. They heard one of the frequent nouns and saw four pictures with referents of the other frequent nouns. Participants could progress to the next phase when they correctly identified each frequent noun four times. After three incorrect responses for a certain noun, the experimenter would provide feedback for the correct answer. When a particular noun had not been identified correctly four times, it would continuously reoccur in the vocabulary test, until four correct answers were reached for that noun as well, in order to make sure all participants had the same starting point when beginning to learn the rule. The length of the vocabulary training could thus vary across participants, but this would not have consequences for the amount of input of the target rule, as the vocabulary training did not expose participants to the functional elements.

After the vocabulary training, participants were exposed to three rule training phases in which they received input on the basis of which they could learn the grammatical rule. During each phase, 40 sentences were presented, adding up to a total of 120 sentences. Each sentence was accompanied by a picture showing the meaning of the sentence. A subset of 108 sentences consisted of a proper name, a verb, a marker and a noun. In these training sentences, the proper name functioned as subject, while the noun was the object. These sentences served as the input to learn the plural marking rule. The other 12 sentences were created using two proper names, a conjunction and a verb. These sentences did not contain a common noun and thus did not provide evidence of the grammatical rule that had to be learned. These sentences were used to create the attention task (see below) and must be regarded as fillers. Examples of all sentence types and their accompanying pictures can be found in Table 2.2. Sentences in the language were always semantically plausible. After rule training phases one and two, participants were given a six-item vocabulary test (one item for each frequent noun), using the same procedure as described earlier. These vocabulary tests were inserted to maintain participants' attention. Participants would receive a sticker after each vocabulary test, regardless of their results.

Table 2.2. Examples of artificial language training sentences and rough translations.

	Example	Translation	Picture
***Tra* + single referent**	*Maria rigarda tra zambo*	'Maria eats (an) apple(s)'	
***Tra* + multiple referents**	*Carlo estima tra pano*	'Carlo looks at (a) cat(s)'	
***Pli* + multiple referents**	*Julia pentura pli anso*	'Julia takes a picture of dresses'	

Filler	*Marco ut Maria pentura*	'Marco and Maria take a picture of each other'	

Note: For sentences with *tra* the noun could be translated both as a singular and a plural. The correct interpretation should be inferred from the visual context.

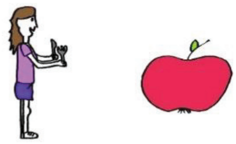
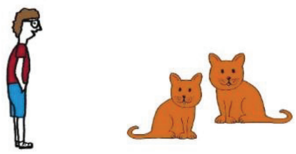
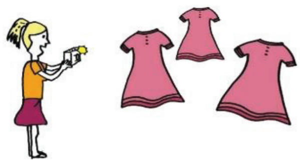
Furthermore, because previous studies indicate attention plays a critical role in statistical learning (Toro, Sinnett, & Soto-Faraco, 2005; Arciuli, 2017), we wanted to know whether the inclusion of a task to maintain attention would lead to higher learning rates. 24 of the 50 participants performed this attention task. In the attention task condition, the four protagonists sometimes could not hear what had been said in the new language correctly. This was indicated by a questioning face of a protagonist after a certain stimulus. When participants saw this face, they had to repeat the previously heard stimulus. This attention task was inserted to keep participants focused; they did not have to repeat what had been said correctly. Participants were introduced to this attention task during the vocabulary training where they had to repeat each noun once. During the rule training phase, they had to repeat four filler sentences per training phase. Filler sentences did not contain a grammatical marker and a noun, and occurred at fixed moments during each phase. Participants who did not perform the attention task also heard these filler sentences, but simply did not have to repeat them. There was no difference in input for the rule, nor salience of this input, between the two conditions.

2.3.2.3 Characteristics of the input

From the 108 training sentences that participants could use to learn the rule during the rule training phases, 72 sentences contained the default marker *tra*, while 36 sentences contained the marker *pli*, which was indicative of number. In half of these 72 *tra* sentences, the noun had one referent. In the other half, the noun had multiple referents. In half of the sentences that contained a noun with multiple referents, two referents were shown, in the other half of the sentences that contained multiple referents, three referents were shown. This distribution was the same for sentences with *pli* that showed multiple referents and *tra* that showed multiple referents. As a result, 36 pictures showed one referent, 36 pictures showed two referents, and 36 pictures showed three referents. Furthermore, the probability that the noun following *pli* had multiple referents was 1. For a noun following *tra*, this probability was 0.5, as the probability that it referred to a single referent was 0.5 as well. Vice versa, when a single referent was shown, the probability of hearing *tra* before the noun was also 1. If a participant saw multiple referents, the probability of hearing *tra* before the noun was 0.5, as was the probability of hearing *pli* before the noun.

Every frequent noun occurred twelve times in the input. It occurred four times with *pli* and eight times with *tra*. When a frequent noun occurred with *tra*, it referred to a single referent four times and multiple referents the other four times. Every infrequent noun occurred three times in the input, once with *pli* and twice with *tra*, of which it once referred to a single referent and once to multiple referents. Each noun referred to one, two or three referents equally often. Every noun occurred equally often over each of the three rule training phases, as did every grammatical marker. An overview of the characteristics of the input can be found in Table 2.3.

Table 2.3. Different types of sentence in the input and how often they occurred per block and with each noun type.

Structure	Total occurrences	Occurrences per noun type	Picture
***Tra* + single referent**	36 times 12 times per block	4 times per frequent noun 1 time per infrequent noun	
***Tra* + multiple referents**	36 times 12 times per block	4 times per frequent noun 1 time per infrequent noun	
***Pli* + multiple referents**	36 times 12 times per block	4 times per frequent noun 1 time per infrequent noun	

Filler	12 times 4 times per block	Did not contain a noun	

2.3.2.4 Test phase

Immediately after the rule training phase, participants took part in a picture matching task to determine whether they became sensitive to the grammatical cue. In this task, participants heard 36 sentences based on the twelve infrequent nouns from the rule training phase, and had to choose which of two pictures matched each sentence. Infrequent rather than novel nouns were used, so participants would feel a target response could be based on what they had been exposed to. Young children reportedly have difficulties with tests in which they have to make decisions that are unrealistic to them (for a methodological review, see Pinto & Zuckerman, 2019), which would be the case when test items contained novel nouns and pictures. We tried to circumvent this by using infrequent rather than novel nouns in the picture matching task. This way, participants could base their response on what they had been exposed to; both pictures would be realistic options, as they would have seen them. Yet, by using infrequent nouns, we could still maximize the chance these children used rule knowledge when giving an answer, as chances are slim that they learned the meaning of these nouns from only three occurrences in the input.

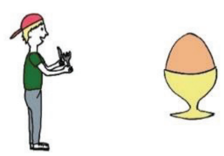
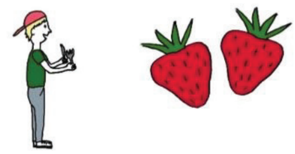
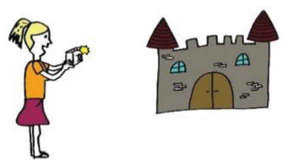
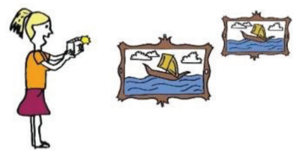
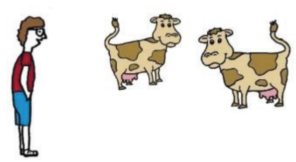
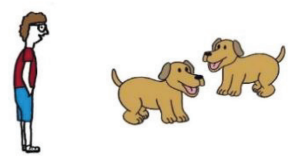
24 out of 36 sentences used during the test phase were experimental items. For experimental items, participants had to choose between pictures with

either one or multiple referents. The two pictures always referred to different referents. Twelve experimental items contained *pli*. For these items, the target picture always showed multiple referents. Another twelve experimental items contained *tra*. Any number of referents would be grammatical for these items. For half of the experimental items containing *tra*, the target picture showed multiple referents. For the other half of these items, the target picture showed a single referent. Every noun occurred once with *tra* and once with *pli* during the test. If participants learned the meaning of the infrequent noun, they should produce a target answer in both conditions. The performance on the sentences with *tra* can be used as a baseline to which we can compare performance on the sentences with *pli*. We hypothesized that sensitivity to the statistical regularity in the input would lead to more target answers on trials with *pli* than on trials with *tra*, because for trials with *pli* participants could base their response on the number of referents the picture showed, whereas the number of referents was not indicative for trials with *tra*.

In addition to the 24 test items, 12 filler items were included. Fillers contained *pli* or *tra*, but showed the same number of referents on both pictures. These fillers were included to avoid that participants would link the grammatical markers to number during the test phase only, because they always had to choose between a picture with multiple and a picture with a single possible referent. Examples of the different types of items can be seen in Table 2.4. Items of the different types were presented in a counterbalanced semi-randomized order.

After the picture matching task, a debriefing took place in which we examined whether participants were aware of their knowledge in a way they could verbalize. During this debriefing, participants were asked how they knew what the correct answer during the picture matching task was, and whether they knew the meaning of the words *pli* and *tra*.

Table 2.4. Examples of test items and their rough translations.

	Marco rigarda pli bovo - 'Marco eats strawberries'	
Pli		
	Julia pentura tra nutro – 'Julia takes a picture of (a) painting(s)	
Tra		
	Carlo estima tra misto – 'Carlo looks at (a) dog(s)	
Filler		

Note: For sentences with *tra* the noun could be translated both as a singular and a plural. The correct interpretation should be inferred from the visual context. In all three examples, the right picture was the target.

2.3.3 Procedure

All stimuli were recorded by a female native speaker of Dutch. The test was administered in a quiet room at the participants' school. The task was presented on a laptop using E-prime (Psychology Software Tools, 2012). During vocabulary training, nouns and their accompanying pictures were presented for three seconds, before automatically moving to the next noun. During rule training, sentences and their accompanying pictures were presented for four seconds, before automatically moving to the next sentence. During the test phase, the experimenter pressed a button on the keyboard that corresponded to the answer the participant gave (z for the left, m for the right picture). Scores from this phase were registered automatically. Both the participant and the experimenter listened to the audio using headphones. The vocabulary training lasted 8 minutes on average, the rule training phase 15 minutes, and the test phase 5 minutes. The full experiment took approximately 30 minutes per participant. Ethical approval for this study was obtained from the University of Amsterdam and passive consent was obtained from children's parents or legal guardians before the start of the study (i.e. children were allowed to participate, unless their caregivers objected to participation).

2.4 Results

To determine whether participants grasped the target regularity and whether the attention task had an effect on task performance, a generalized linear regression model with mixed effects and orthogonal sum-to-zero coding was carried out. This analysis was carried out in R (R Core Team, 2015) using the lme4 package (Bates, Maechler, Bolker & Walker, 2015) where needed. The generalized linear regression model took the responses from the picture matching task (1 or 0) as a dependent variable, marker type as a within-participants fixed effect, attention task as between-participant fixed effect, participant as a between-participants random effect and item as a within-participants random effect. We did not include any random slopes. Our fixed effects were included in this model, because we were *a priori* interested in their contribution to the outcome (Gelman & Hill, 2007). Orthogonal sum-to-zero contrast coding was applied to our binary fixed effects (i.e., marker type and attention task; Baguley, 2012, p590-621). As we aimed to keep the model as fully specified as possible (Barr, Levy, Scheepers & Tily, 2013), we increased the number of possible iterations to 100 000 (Powell, 2009) to solve issues of non-convergence. This enabled us to report on the maximal random effect structure justified by our data (Jaeger, 2009). We report simple rather than standardized effect sizes (Baguley, 2009) and Wald confidence intervals (Agresti & Coull, 1998).

The descriptive statistics can be found in Figure 2.2 and Table 2.5. Results from the generalized linear regression model showed that participants gave more target answers when sentences contained *pli* than when sentences contained *tra* (OR = 1.509, 95% CI = [1.040, 2.190], z = 2.167, p = .030), suggesting they were sensitive to the statistical regularity in the input.[1] Results did not show a significant main effect of attention task (OR = 0.847, 95% CI = [0.647, 1.109], z = -1.209, p = .227). The interaction between performing the attention task and marker approached significance (OR = 1.557, 95% CI = [0.973, 2.492], z = 1.845, p = .065).

[1] Possibly, this response pattern was due to a bias for pictures showing multiple objects. However, we found that for sentences with *tra,* where half of the items had a singular target, and half of the sentences a plural target, the correct answers were evenly distributed over these two target types. Thus, there is no reason to assume children had a plural bias when providing answers in the picture matching task. The results of this analysis can be found in the supplementary materials on our OSF page.

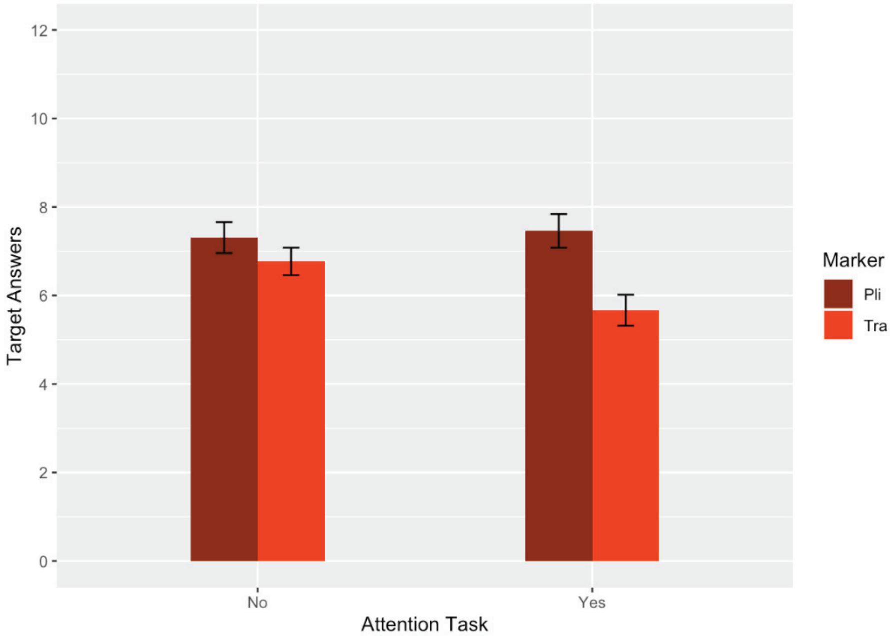


Figure 2.2. Graph depicting the results from the picture matching task. Scores indicated the number of target answers produced and could range from 0-12.

Table 2.5. Scores from the picture matching task indicating the number of target answers produced.

	With attention task (n = 24)			**Without attention task (n = 26)**			**Combined (n=50)**		
	M	**SD**	**Range**	**M**	**SD**	**Range**	**M**	**SD**	**Range**
Pli	7.46	1.86	4-11	7.31	1.76	5-12	7.38	1.79	4-12
Tra	5.67	1.71	3-10	6.77	1.58	4-10	6.24	1.72	3-10

Note: Sentences with grammatical marker *pli* were predictable. Scores could range from 0-12.

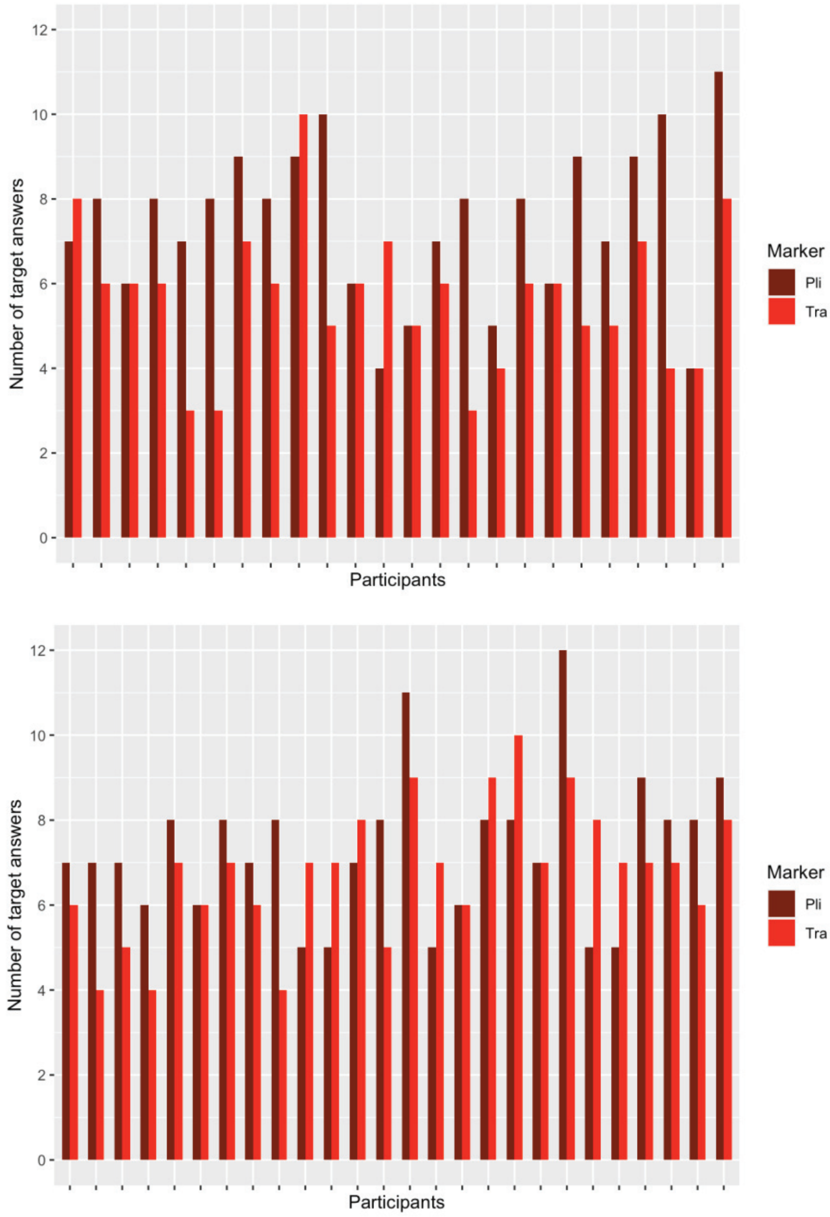


Figure 2.3. Graphs depicting the number of target answers produced during the picture matching task per participant. Above are scores from participants who performed the attention task, below those from participant who did not perform this task. Scores could range from 0-12.

To shed further light on children's individual differences in learning the grammatical marker, we calculated a learning score for each child. As input for this score we used the number of target answers for sentences with *pli* and sentences with *tra* of every individual participant, which can be observed in Figure 2.3. We calculated the learning score by subtracting the number of target answers on sentences with *tra* from the number of target answers on sentences with *pli*. A larger difference in a positive direction indicates that the participant gave more target answers on sentences with *pli* than on sentences with *tra*. Difference scores per participant for the picture matching task can be seen in Figure 2.4. In total, 62% of the participants (N = 31) exhibited a positive difference score, 16% of the participants (N = 8) a neutral difference score, and 22% of the participants (N = 11) a negative difference score.

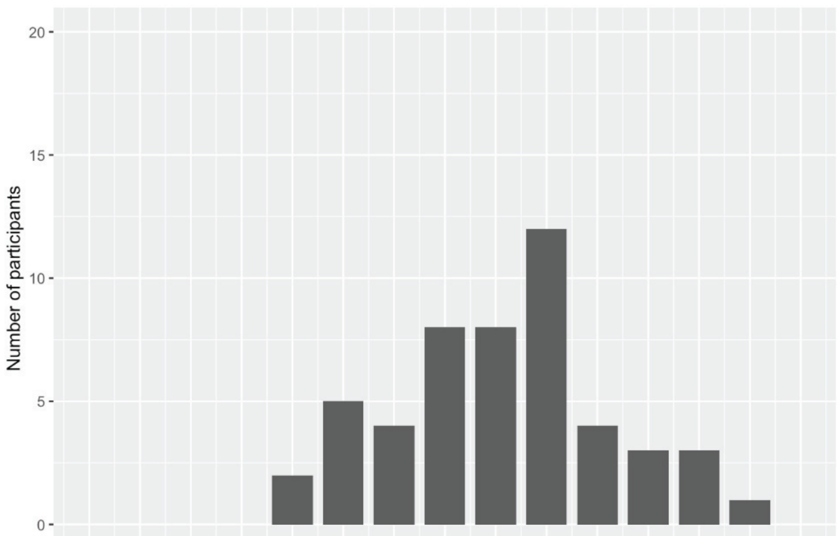


Figure 2.4. Graph depicting the difference scores from the picture matching task per participant. A positive score indicates a participant gave more correct answers for sentences with *pli* than sentences with *tra,* and thus indicates learning of the regularity.

During the exit interviews, none of the participants was able to report on the regularity. When asked how they came to a decision on the picture matching task, participants either reported they did not know how they made a decision (74%, N = 37), or they claimed that they had heard the sentences before and remembered what they meant (26%, N = 13). When asked for the meaning of *pli* and *tra*, children did not know the meaning of these words (80%, N = 40), provided the meaning of noun that they were exposed to (18%, N = 9) or came up with a completely new meaning (2%, N = 1).

2.5 Discussion

The results from this cross-situational statistical learning experiment show that kindergarteners were able to detect a regularity of a number feature encoded by a grammatical marker. The distributional properties in the input were the only cue to detect the regularity, and we hypothesized that statistical learning mechanisms facilitate acquiring this grammatical marker. Previous studies have shown that statistical learning plays a role in word segmentation, word referent learning and the acquisition of semantic categories. Although the statistical information that learners needed to use varied considerably between these studies, learners in all cases had to associate particular pieces of information with each other. Building on these studies, we hypothesized that a similar learning mechanism is involved in acquiring grammatical markers, such as markers of nominal number.

The present study tested whether kindergarteners could grasp a statistical dependency between an auditory linguistic element and a visual plural interpretation of this marker. Importantly, we investigated this in a miniature language learning experiment, using input that resembled natural linguistic input more closely than earlier studies that investigated similar grammatical structures (e.g. Raviv & Arnon, 2018; Tal & Arnon, 2020). Indeed, the results are compatible with the idea that a domain general capacity for statistical learning plays a role in learning such meaningful grammatical elements. However, we should note that, although this learning effect was significant, it was relatively small. This could mean that while statistical learning is involved in detecting the distributive properties of a grammatical element, it interacts with other learning mechanisms that lead to acquisition of the meaning (e.g. Lidz & Gagliardi, 2015; Yang & Montrul, 2017).

Results also indicated that learning this type of regularity slightly improved when participants performed a task intended to increase their attention. However, evidence for this is far from conclusive as the effect only approached significance. If this finding were to hold, this would be in line with earlier studies that indicated that statistical learning might benefit from extra attention (e.g. Toro et al., 2005; Arciuli, 2017). The lack of conclusive outcomes with regard to the role of attention in our study could be a result of the small main effect of learning we found. Alternatively, it might show attention plays some role in statistical learning, but that its role is relatively small.

Several other factors may have influenced how well the marker was learned, as we observed considerable individual differences in learning scores. For some children, the difference between *pli*-and *tra*-sentences was quite large, even though they did not score very well overall. Some children scored highly on *pli*-items, which suggests learning of the marker, but also scored well on *tra*-items. All kinds of factors not measured in this study could have caused such behavioral differences. For example, children may have varied in how easily they picked up the meanings of words, there may have been effects of knowing or speaking other languages besides Dutch, or children could have varied in general cognitive capacities. The goal of the current study was to investigate whether children could learn this marker, and not which (external) factors make some children better learners than others. Therefore, we chose not to collect any data about the (non-)linguistic background of these children, to keep the sample representative of the population of interest (Kruskal & Mosteller, 1979; Kukull & Ganguli, 2012 for discussions about different types of representation). This choice, however, does not enable us to explore if and how particular individual characteristics might be related to the observed individual variation in our sample.

Furthermore, responses from the exit interviews also seem to be in line with the idea that statistical learning happens without any awareness involved (e.g. Reber, 1967; Aslin et al., 1998; Baker, et al., 2004; Arciuli & Simpson, 2012; Kidd, 2011; Rebuschat, 2013). Not a single child could report on the plural marking rule that was present in the input. However, these results should be interpreted with some caution, as awareness might exist at different levels. Recall from the discussion in Chapter 1 that learners may possess phenomenal awareness (when they can verbalize their experiences) or could be aware at the level of access (when they cannot verbalize or remember their experiences coherently; Cleeremans, 2008, 2011, 2014). A verbal report as used in this experiment taps only into the former type of awareness (Timmermans & Cleermans, 2015) and is thus inconclusive about the involvement of awareness of the latter type. Perhaps, participants in this experiment were aware of the regularity in a way they could not verbalize. To establish whether such awareness is involved in statistical learning a novel kind of experimental method should be used. Chapter 3 discusses how such an experiment could look like, and in Chapter 4 this method is combined with the miniature language that was presented in this chapter.

In sum, this study provides evidence that statistical learning might support the acquisition of a meaningful grammatical element. However, it is important to note that participants in this experiment had to learn a grammatical marker with a relatively simple abstract meaning. Natural languages also exhibit more complex abstract grammatical patterns than the rule that was presented in our artificial language. Whether statistical learning mechanisms can account for the acquisition of more complex patterns remains unknown, let alone more complex patterns in the real world. Further research is needed to gain a better understanding of the role of statistical learning in language acquisition, especially because it is difficult to determine to what extent statistical learning studies in the lab scale up to naturalistic language learning environments. If more complex patterns can be acquired through statistical learning as well, this would give further reason to assume that this learning mechanism could be able to support a broad array of language acquisition processes. The study presented in this chapter is a small, but important step in that direction.

Chapter 3

The opt out paradigm*

In Chapter 1, we described how it is commonly assumed that children acquire language implicitly without any awareness of form and grammar (e.g. Wijnen, 2013). We also established that very little research has been conducted to investigate this assumption. In contrast, the role of awareness in adult second language acquisition has been extensively investigated, and there is a wide range of theoretical perspectives and experimental methods that address the issue (Hulstijn, 2015; Andringa & Rebuschat, 2015; Rebuschat, 2015). Unfortunately, the work done in second language acquisition (SLA) does not yield suitable methodological tools for assessing meta-linguistic awareness in children. Here, we aim to bridge this methodological gap by applying a paradigm that has been used to investigate awareness in animals (Hampton, 2001, 2009) to a linguistic context.

Before we introduce this paradigm, we must clarify our conception of awareness. As briefly introduced in Chapter 1, we follow the idea that learners may either be phenomenally aware (P-awareness) or aware at the level of access (A-awareness; Cleeremans, 2008). P-awareness refers to the situation in which a person has a subjective experience of the object of awareness. Someone can be called A-aware when an experience is accessible to one's awareness, but this cannot be expressed subjectively. Several characteristics are typically associated with P-awareness; one should be able to verbalize and remember a particular

---

* This chapter is a slightly modified version of a published article: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (2019). The opt out paradigm: First steps towards a new experimental method that measures meta-linguistic awareness. *Dutch Journal of Applied Linguistics, 8*(2), 206-227.

experience one is aware of as a coherent whole. One can only be aware at this phenomenal level if one's mental representation of the experience is strong enough. This type of awareness bears resemblance to what is often called explicit knowledge. However, verbalization, coherence and memory might not be necessary requirements for awareness (Allport, 1988; James, 1890; Dennett, 1993). Less strong mental representations could be accessible to awareness, even though one cannot verbalize or memorize these representations coherently. This is perhaps the type of awareness that can be ascribed to animals who are clearly unable to verbalize their experiences, but nevertheless are shown to be aware of such experiences (see de Waal 2016 for an overview, although not in terms of A- and P-awareness). In this sense, A-awareness should be distinguished from implicit knowledge, which is knowledge someone has no awareness of whatsoever. Instead, A-awareness could be characterized as a state of awareness between explicit and implicit knowledge.

If we apply these forms of awareness to the study of language, someone can be said to be phenomenally aware, when s/he is able to explicitly describe certain linguistic patterns in the input. On the other hand, for instance, one can be aware at an access level if one perceives that someone is speaking with an accent, and one can remark that 'something' is going on, without being able to determine which accent is used, or that an accent is used at all. This A-awareness resembles what has been called *noticing* in the L2 literature (Schmidt, 1990). Furthermore, the object of meta-linguistic awareness could consist of different linguistic phenomena. Someone could be aware *that* s/he is learning a language, but here we are not interested in this holistic type of meta-linguistic awareness. With meta-linguistic awareness we imply awareness of *what* one is learning in a particular language.

It is commonly assumed that language learners are only aware of linguistic patterns in the phenomenal sense if they can verbalize and remember them: meta-linguistic P-awareness. However, learners in general and children in particular might possess awareness of the patterns they acquire, without being able to verbalize those patterns: meta-linguistic A-awareness. Traditional methods of investigating meta-linguistic awareness, such as verbal reports or confidence judgement tasks, typically measure P-awareness (Timmermans & Cleeremans, 2015; Batterink et al., 2015) and thus seem unsuitable to investigate awareness in children. Therefore, we adopt the 'opt out paradigm', which has been used to investigate (A-)awareness in animals (Hampton, 2001, 2009). The opt out paradigm assesses awareness without requiring learners to verbalize their experiences. In this paradigm participants show strategic behavior that reveals their awareness. We believe this method can be used to assess whether children are aware of the linguistic patterns they are acquiring.

In this chapter, we explore the possibility of using the opt out paradigm as a measure of meta-linguistic awareness. As this is a novel method with regard to children, the aim of the study presented here was twofold. First, we wanted to see whether the opt out experiment can reveal awareness of acquired knowledge in adults. In addition, we wanted to determine whether the procedure of our experiment is suitable for children. A positive answer to both of these questions was a requirement for the study we present in Chapter 4, where we investigate the role of awareness in child language acquisition using an adaptation of this experiment. In the remainder of the present chapter, however, we will elaborate briefly on the role of awareness in language acquisition, and why it is worthwhile to develop a method for gauging awareness in child language learning. Subsequently, we will discuss in more detail how awareness is typically investigated and how our method can contribute to the range of existing methods.

3.1 Awareness in language acquisition

The role of meta-linguistic awareness in the process of acquiring a new language has been widely investigated (e.g., Hulstijn, 2015; Andringa & Rebuschat, 2015). Some authors have argued that awareness of the grammatical structures in a second language is necessary to acquire those structures (Schmidt, 1990; DeKeyser, 2003), whereas others have claimed awareness merely facilitates the acquisition process (Ellis, 2003). Others have suggested that the presence of awareness is circumstantial and a consequence of how the language is learned, but has no influence on the implicit linguistic knowledge that is gained (Krashen, 1981).

Theories of child language acquisition are often less explicit about the role of meta-linguistic awareness in the acquisition process. Many theories assume that meta-linguistic awareness hardly plays any role. This seems to be the case in usage-based theories (Tomassello, 2000), in generative approaches (Chomsky, 1986; Radford, 2004) and also in theories that assume language learning to be a statistical process (Aslin et al., 1998; Endress & Bonatti, 2007; Erickson & Thiessen, 2015; Saffran et al., 1996). Whether children map forms onto functions, set grammatical parameters or track dependencies between linguistic elements, it is generally assumed that children are unaware that they are carrying out these cognitive processes and that they are unaware of the regularities they acquire as a result.

The idea that primary language learning occurs without awareness also emerged in frameworks that seek a neurological foundation for language, such as the declarative/procedural memory model (Ullman, 2001, 2004). In this model, procedural memory is responsible for pattern recognition in the auditory as well as the visual domain (Packard, 2009). Procedural learning is often equated with implicit learning, as people typically are assumed to be unaware they are acquiring grammatical regularities and to be unable to reflect on their acquired regularities (West, Vadillo, Shanks & Hulme, 2018). On the other hand, declarative memory is thought to handle the storage of more arbitrary information, concrete facts and events. Declarative knowledge is said to be gained more explicitly, because we learn this kind of linguistic information consciously and we are able to retrieve this knowledge (Ullman, 2016). Many studies have shown that domain-general procedural memory is related to the acquisition of linguistic patterns and that child language acquisition in particular might rest on this memory system (Evans et al., 2009; Hsu, Tomblin & Christiansen, 2014; Lum, Conti-Ramsden, Page & Ullman, 2012; Misyak & Christiansen, 2012; Misyak, Christiansen & Tomblin, 2010; Ullman & Pierpont, 2005). Because children are often portrayed as more skillful procedural learners, they are assumed to be unaware of the linguistic patterns they acquire (Wijnen, 2013).

However, whether children indeed acquire language completely implicitly is an issue that has hardly been put to test empirically. Importantly, mapping form unto function, setting parameters, tracking statistical regularities or proceduralizing knowledge might play a role in language acquisition, but these processes need not necessarily be dissociated from awareness. One reason why awareness during acquisition has not been investigated in children may be that people tend to equate awareness with P-awareness. Possibly, children do not have this type of awareness of the language they acquire, because they are usually not able to verbalize the patterns of language. Yet, children could still possess A-awareness when acquiring a language: they are aware of the linguistic patterns they acquire though they are unable to verbalize this awareness. It is conceivable that some well described phenomena within language acquisition, such as overgeneralization (Barac & Bialystok, 2012), stem from children's A-awareness, but to our knowledge, links between such phenomena and awareness in children have not been as systematically investigated as they have in the field of second language acquisition. If we want to make claims about the role of (A-)awareness in child language acquisition, the question to answer is thus how such awareness can be measured.

3.2 Measuring Awareness

Typically, when researchers investigate the role of awareness, direct subjective methods are used (Timmermans & Cleeremans, 2015). These methods often consist of two parts. First, experimenters assess whether a participant learned or acquired a certain rule or regularity. Afterwards, the experimenters investigate whether a participant was aware of what s/he has learned or acquired.

To measure learning, experimenters can use grammaticality judgement tasks or Wug tests, obtain event-related potential (ERP) signals, gather eye tracking data, or measure reaction times. Crucially, none of these measures reveal anything about the awareness that learner have about their gained knowledge. If reaction times show learners have acquired a certain regularity, they are not necessarily aware of this regularity. Their behavior can be based entirely on implicit knowledge. For grammaticality judgment tasks and Wug tests (Berko, 1958), the same applies, although some scholars have argued they are a measure of meta-linguistic awareness (Bialystok, 1986; Barac & Bialystok, 2012). Perhaps some level of meta-linguistic awareness is needed to perform these tasks, as one has to recognize regularities in language or identify correct and incorrect sentences. The point is, however, that grammaticality judgment tasks and the Wug test do not reveal the status of the knowledge that was used to make choices in the tasks. In such tasks, learners may make choices because they are aware of *what* the regularities are, or their behavior could also be the result of implicit knowledge of these regularities.

A similar rationale applies to the research showing that babies are able to distinguish between their native language and a foreign language (e.g. Bosch

& Sebastián-Gallés, 2003) or research indicating that infants infer that non-native languages serve communicative purposes as well (e.g. Vouloumanos, 2018). In such studies infants could, for example by a preferential looking paradigm, show signs of awareness *that* language serves particular functions and *that* there is a difference between languages, but this paradigm does not reveal whether infants are aware of *what* these functions are or *what* the differences are. Of course, infants could possess such awareness, but these methods are not suitable to measure it.

Measures like grammaticality judgement tasks, Wug tests and preferential looking paradigms can thus chart learning, but they cannot gauge whether learners developed awareness of *what* the regularities in the acquired language are. To tap into this sort of awareness, learners typically have to report verbally on whether they were aware of the regularity (e.g. Batterink et al., 2015) or partake in a confidence judgement task, indicating how certain they are about their grammatical judgment (e.g. Bertels et al., 2015). These methods directly ask participants to reflect on the object of awareness in a way that relies on their subjective experiences, as they have to verbalize their knowledge or indicate how confident they are. In other words, they tap into P-awareness.

If meta-linguistic awareness exists at the A-level, a different kind of measurement is required. To measure A-awareness, indirect objective methods are required (Timmermans & Cleeremans, 2015). In such methods, experimenters observe behavior that indirectly reflects some form of awareness, but does not require any subjective introspection, verbalization or reflection on knowledge. This makes these methods specifically suitable to tap into A-awareness and they may be a more fruitful way to investigate the role of awareness in child language learning.

An example of an indirect objective method that can potentially be adapted to a linguistic context is the opt out paradigm, which is commonly used in studies with non-human animals (de Waal, 2016; Hampton, 2001, 2009).

Unlike humans, non-human animals cannot verbally report on their awareness. This method is thus particularly interesting in this respect. Experiments reported in Hampton (2001) rely on the fact that monkeys like peanuts more than pellets. Rhesus monkeys were trained to remember a picture of a flower. After a certain time, the monkeys were shown an array of pictures of flowers and had to identify the flower they had seen before. If they identified the right flower they received a peanut, which they really like. If they answered incorrectly, they would not receive anything. After a training period, a new condition was introduced which gave the monkeys the opportunity to opt out of the test and for which they would receive a pellet. While the monkeys prefer peanuts over pellets, opting out would give them a guarantee of receiving at least some food. In the experiment, the option to opt out was given either shortly after seeing the flower they had to remember or after a longer period of time.

During the experiment the monkeys chose to opt out more often when this interval was longer and thus their memory of the target flower had decayed more. Hampton argued that this is indicative of at least some level of awareness on behalf of the monkeys. When the interval was longer, it would be a better strategy to opt out, because the chance of guessing the right picture was lower due to forgetting. Thus, the risk of not receiving any reward at all was relatively high. In order for the monkeys to make this type of decision, they must have been at some level aware of the fact that they had a certain memory of the picture of the flower that decays over time. Additionally, they could take this decaying memory into account when they had to make a decision that was based on the status of this memory. The monkeys thus did not have to engage in some subjective introspection with regard to their memory, but instead performed some objective observable behavior that was the indirect effect of their awareness. As such, this opt out paradigm is an indirect, objective measure that taps into A-awareness.

The main goal of this chapter is to investigate whether the opt out paradigm can be adapted to a linguistic context successfully, such that it can be used in research, such as presented in Chapter 4, to investigate the role of awareness in child language acquisition. As this chapter only presents a first step towards implementing this paradigm, the concrete goals of the here presented study were twofold. To start off, we wanted to see whether the opt out experiment could reveal awareness of acquired knowledge in adults. Next, we wanted to know whether the procedure of our opt out experiment could be suitable for children. Several studies show that different populations react differently to rewards (Frederick, 2005; Eigsti et al., 2006), with for example children being much more likely to go for certain rewards than to seize the opportunity to obtain a bigger reward after a period of time or after the performance of a task (Mischel, Shoda & Rodriguez, 1989). Given this observation, we also wanted to determine that the reward for opting out in our experiment is not too appealing, or that the task participants had to perform to earn a bigger reward is not too difficult. If either is the case, children in our experiment will be biased to opt out, regardless of their level of awareness. However, if we find that children choose to opt in and out, then this procedure should be suitable to probe into children's awareness.

3.3 Method

3.3.1 Participants

Twenty-six Dutch speaking adults (8 males, 18 females, *M* = 22;9, *SD* = 4;4) took part in the opt out experiment. No restrictions were imposed upon taking part in this pilot. A relatively large number of students of linguistics were part of the adult sample. Additionally, 48 Dutch speaking children (26 males, 22 females, *M* = 5;3 *SD* = 0;9) took part in the experiment. All children were in kindergarten and did not have any diagnosed language or communication disorders. Ethical approval for this study was obtained from the University of Amsterdam and before children participated, active consent was obtained from their parents or legal guardians.

3.3.2 Opt out experiment

The experiment consisted of a learning phase and an opt out phase. During the learning phase, participants had to discover a correspondence rule based on the Dutch determiner system. In Dutch, nouns can be of common or neuter gender, which each have their own definite article; *de* for common and *het* for neuter nouns. (*de koe* (‘the cow’, common) with *het paard* (‘the horse’, neuter). Participants were shown pictures that were accompanied by Dutch sentences in which a noun was replaced by a nonsense word. The sentences with a nonsense word would still refer to the concepts depicted in the illustration, but participants could arrive at the meaning of the nonsense words through the correspondence between the determiner that preceded the nonsense word and the determiner of the word that was illustrated in the picture. For example, the Dutch word *koe* ´cow´ was replaced by the nonsense word *lino* which was always preceded by the determiner of the Dutch word for cow (‘de’). This resulted in the noun phrase *de lino*, which referred to a cow. In contrast, the Dutch word *paard* ´horse´ was replaced by the nonsense word *orbo* which was always preceded by the determiner of the Dutch word for horse (‘het’). This resulted in the noun phrase *het orbo*, which referred to a horse.

To learn the correspondence rule, participants were presented with a spoken sentence and were shown two pictures. The sentence described only one of the two pictures and participants had to decide which picture matched the sentence they heard. The input during the learning phase consisted of 36 sentences of three different types; 12 sentences with a definite article and a Dutch noun (3a), 12 sentences with a definite article and a nonsense word (3b), and 12 sentences with an indefinite article and a nonsense word (3c). Figure 3.1 shows an example of the pictures that participants had to choose between. All test items can be found in the Appendix.

(3a)	In	het	gras	staat	de	koe
	In	the	grass	stands	the	cow
	'The cow is standing in the grass.'					
(3b)	In	het	gras	staat	het	orbo
	In	the	grass	stands	the	orbo
	'The orbo is standing in the grass.'					
(3c)	In	het	gras	staat	een	orbo
	In	the	grass	stands	a	orbo
	'An orbo is standing in the grass.					

Figure 3.1. The two pictures between which a participant had to choose, when s/he heard the sentence *In het gras staat de koe* ('The cow is standing in the grass'). The right picture was the target.

The pictures always depicted two possible meanings of the sentence. One depicted a concept that in Dutch is described by a common noun and the other by a neuter noun (see Figure 1 for an example showing a cow and a horse). A correct decision in the picture matching task could be made on the basis of the final noun phrase, as the rest of the sentence could refer to both pictures. In the case of sentences with a Dutch noun in the final noun phrase, this decision could be made on the basis of the lexical meaning of that noun phrase. In the case of sentences that ended with a definite article and a nonsense word, this decision could be made on the basis of the earlier mentioned correspondence rule.

For the sentences ending with an indefinite article and a nonsense word, a correct decision could only be made if participants remembered the meaning of the nonsense word. In Dutch, gender is only marked on the definite article, but not on the indefinite article: both common and neuter nouns receive the article *een* 'a'. Therefore, the indefinite article introducing a nonsense words was not predictive of its meaning. Participants could give a correct answer in these cases, when they matched a nonsense word to its meaning when the correspondence rule could be used. They then had to remember this meaning and correctly apply it to the nonsense word when the correspondence rule could not be used to make the right decision.

Test items were presented semi-randomized, so that the Dutch nouns were always presented before the definite nonsense equivalent was presented. This way participants were presented with the Dutch noun and its article before the nonsense equivalent was presented, such that they had the opportunity to apply the correspondence rule to the right Dutch noun. Furthermore, a sentence ending with a definite article and a nonsense word had to be presented before its indefinite equivalent was presented, so (3c) could only be played after (3b) was presented. This was to assure that participants could give a correct answer to the indefinite cases, if they remembered the nonsense words and their matching concepts correctly.

Before this learning phase started, participants were presented with some practice items with both existing Dutch words and nonsense words. Importantly, after they had completed two-thirds of the learning phase, participants received a reward when they made a correct decision. At this point all sentences with an existing Dutch word had been presented, as well as half of the sentences with a definite article and a nonsense word, and half of the sentences with an indefinite article and a nonsense word. For adults, the reward consisted of 2 euros, the currency being Monopoly® money. They were encouraged to obtain as much money as possible. Participants did not keep their reward after the experiment. Children helped a rabbit cross a river with 10 stepping stones to obtain a carrot. The rabbit would move 2 stones for every correct answer. Children received a sticker for every carrot the rabbit obtained.

In the opt out phase, participants heard the same sentences as before, up to and including the determiner, but without the last word, while seeing a blank screen. Participants would hear sentences as in (4).

(4a) In het gras staat de/het …
‘In the grass stands the …’
(4b) In het gras staat een …
‘In the grass stands a …’

They were then given two options. They could hear the whole sentence including the last word. When hearing the full sentence, they would see the two pictures and had to make a decision between these pictures. As before, if they decided correctly they would earn 2 euros/2 steps and if they made a wrong decision, they would not obtain any reward. Participants could also choose to opt out. In this case, they decided not to hear the full sentence, but simply to move on to the next test item. When opting out, participants would receive 1 euro or the rabbit would move 1 stepping stone. Thus, in order to earn as much money or obtain as many carrots as possible, participants had to consider their chances of providing a correct answer. If participants decided to hear the full sentence, the sentence would always end with a nonsense word that occurred in the learning phase.

The opt out phase consisted of only the sentences with nonsense words that were presented in the learning phase. There were 12 items that ended with a definite article that would end with a nonsense word if heard fully, and 12 items that ended with an indefinite article that would also end with a nonsense word if heard fully. In order to make sure that participants understood the consequences of opting for or against hearing the full sentence, they were presented two practice items before the test phase, to familiarize themselves with the procedure of the opt out phase. During the entire experiment, the target answer was counterbalanced between left and right.

3.3.3 Procedure

The experiment was administered to adult participants in a quiet room at the University of Amsterdam by one experimenter. The task was presented on a laptop and scores were recorded by hand using paper and pencil. Afterwards, there was an informal debriefing. Adults were not compensated for their participation. For children, the test was administered in a quiet room at their schools. The task was presented on a laptop using E-prime (Psychology Software Tools, 2012) and scores were registered automatically. Children were allowed to keep the stickers they earned during the experiment. Afterwards a debriefing took place in which children were asked whether they knew what the nonsense words meant, how they knew this, why they decided to opt out and why to opt in.

3.3.4 Analysis

All analyses were carried out in R (R Core Team, 2015) using the lme4 package (Bates, Maechler, Bolker & Walker, 2015) where needed. As mentioned before, the aims of this experiment were different for the two test groups. For adults, we wanted to know whether opting out can be used as a measure of meta-linguistic awareness, once a rule is learned. For children, we wanted to know whether they show sensitivity to a reward system based on linguistic input. We were thus not interested in group differences with regard to learning the correspondence rule. Therefore, separate analyses were carried out for the two groups. For each group, generalized logistic mixed effect models were used to investigate whether participants learned the rule in the learning phase of the experiment and whether they were aware of this in the opt out phase. Pearson's correlations were used to investigate possible relationships between the scores on both parts of this test.

3.4 Results

Results from the learning phase can be found in Table 3.1. To determine whether adults learned the correspondence rule, a generalized logistic model with mixed effects and orthogonal sum-to-zero contrast coding was carried out (see Baguley, 2012, p. 590-621). This model took the target score as a dependent variable, article type as a within-participants fixed effect, participant as a between-participants random effect and item as a within-participants random effect. Results showed a significant main-effect of article type (OR = 2.20, 95% CI = [1.21 ... 4.32], z = 2.340, p = 0.019). Adults gave more correct answers when the nonsense word was introduced by a definite article than when it was introduced by an indefinite article. This suggests that adults picked up on the correspondence rule.

Table 3.1. Scores from the first phase indicating the amount of target answers produced.

	Adults (n = 26)			**Children (n = 48)**		
	M	**SD**	**Range**	**M**	**SD**	**Range**
Definite article	8.12	2.34	3-12	6.10	1.43	4-10
Indefinite article	6.35	1.38	3-9	6.10	1.52	3-9

Note: Sentences with a definite article were predictable. Scores could range from 0-12.

To determine whether children learned the correspondence rule, a generalized logistic model with mixed effects and orthogonal sum-to-zero coding was carried out too. This model took the target score as a dependent variable, article type as a within-participants fixed effect, participant as a between-participants random effect and item as a within-participants random effect. Results did not show a main effect of article type (*OR* = 0.98, 95% CI = [0.59 ... 1.64], z = -0.080, p = 0.936). Children did not give more correct answers when the nonsense word was introduced by a definite article than when it was introduced by an indefinite article. This does not provide any evidence that children picked up the correspondence rule.

Results from the opt out phase are shown in Table 3.2. To determine whether adults not only learned the correspondence rule, but also became aware of their knowledge of it, a generalized logistic model with mixed effects and orthogonal sum-to-zero coding was carried out. This model took the opt out score as a dependent variable, article type as a within-participants fixed effect, participant as a between-participants random effect and item as a within-participants random effect. Results showed a significant main-effect of article (*OR* = 11.07, 95% CI= [1.86 ... 65.92], z = 2.696, p = 0.007). Adults opted out more often when the sentence ended with an indefinite article than when it ended with a definite article. This suggests that adults were aware of the correspondence rule.

To determine whether children became aware of the correspondence rule, a generalized logistic model with mixed effects and orthogonal sum-to-zero coding was carried out. This model took the opt out score as a dependent variable, article type as a within-participants fixed effect, participant as a between-participants random effect and item as a within-participants random effect. Results did not show a main-effect of article (*OR*= 1.00, 95% CI = [0.73 ... 1.38], z = 0.025, p = 0.980). Children did not opt out more often when the sentence ended with an indefinite article than when it ended with a definite article. This does not provide any evidence that children were aware of the correspondence rule. Nevertheless, it is important to note here that each child who completed the experiment opted in and opted out at least once during this phase. This means that each child tried out both possible options at least once during this phase. We can conclude from this that a possible bias towards either of the two options is not so overwhelming that it would lead ceiling effects.

Table 3.2. Scores from the opt out phase indicating the amount of times a participant opted out.

	Adults (*n* = 26)			**Children (*n* = 48)**		
	M	**SD**	**Range**	**M**	**SD**	**Range**
Definite article	1.96	2.73	0-12	6.04	3.75	0-12
Indefinite article	5.54	4.51	0-12	6.04	3.59	0-12

Note: It should be more appealing to opt out in case of sentences with an indefinite article. Scores could range from 0-12.

3.4.1 Correlation

To determine whether a correlation existed between awareness in the opt out phase and learning in the learning phase for the adult participants, a difference score was calculated for both phases. For the learning phase, this score was calculated by subtracting the number of target answers on sentences with an indefinite article from the number of target answers on sentences with a definite article. A larger difference expresses a greater likelihood that the participant had learnt the correspondence rule. For the opt out phase, this score was calculated by subtracting the number of times a participant decided to continue when a sentence ended with an indefinite article from the number of times a participant decided to continue when a sentence ended with a definite article. If this difference score was higher, it was more likely that a participant was aware of the rule s/he had learnt. A Pearson's correlation test showed a correlation between learning and awareness for the adult participants ($r = 0.477$, $p = 0.014$).

3.5 Discussion

We have seen that, during the learning phase, adults gave more correct answers when sentences ended with a definite article than when sentences ended with an indefinite article. This indicates that adults, as a group, most likely detected the correspondence rule that determined that the gender of the Dutch nouns corresponds to the gender of their nonsense equivalents. Only in the case of sentences that ended with a definite article could this rule be applied. Therefore, it was expected that they would score better in this condition.

In the opt out phase, adults opted out more often in the case of sentences with an indefinite article than during sentences with a definite article. If participants were either A- or P-aware of the correspondence rule it would be more rewarding to opt out in the case of sentences ending in an indefinite article, because in such cases, participants could not give a correct answer on the basis of a rule. Therefore, it was advantageous to choose the guaranteed smaller reward. In contrast, if participants were aware of the correspondence rule, it would be more rewarding to hear the full sentence and try to give an answer when sentences ended with a definite article. In these cases, a correct decision could be made on the basis of the correspondence rule. The results thus indicate that participants as a group were probably aware of the correspondence rule.

Finally, a correlation was found between the degree to which the rule was learned and the degree to which participants opted out. This could indicate two things. Either, participants develop awareness of the correspondence rule before they can grasp it, or, awareness is a consequence of grasping the correspondence rule. Due to the nature of our experiment we cannot conclude anything about the causality within this relationship.

Our results do not provide evidence that children detected the correspondence rule. Therefore, it is not surprising that our results do not provide evidence that children were aware of the correspondence rule during the opt out phase. Recall, however, that this particular experiment was not carried out to make any inferences about the development of meta-linguistic awareness in children. Instead, we wanted to know whether the procedure of our experiment was suitable for children. More concretely, we wanted to know whether the task was feasible for children, and that the reward for opting out in our experiment was not too appealing. The results do not give reasons to assume the method is faulty. Although a proper linguistic motivation to choose one option over the other lacked in the current experiment, every individual child opted in and opted out at least once during the test. This indicates that our procedure did not lead to an overwhelming opt out bias. Research like the study in Chapter 4 can tell us whether or not children are able to opt out strategically on the basis of a regularity they can acquire. If they are able to do so, this could be indicative of A-awareness of the linguistic pattern they have acquired.

The main goal of our experiment was to investigate whether the procedure of our opt out experiment could be used in other research to measure meta-linguistic awareness in children. For this to work, participants need to be sensitive to a reward system that is based on linguistic input, as this is a prerequisite for successfully adapting this paradigm originally developed by Hampton (2001, 2009). The findings suggest that the opt out paradigm can indeed be adapted to a linguistic context, because adults showed strategic behavior based on their rule knowledge, and because our procedure seemed suitable for children, as there was no clear opt out bias in this population. Nevertheless, some issues should be taken into consideration before this method can be implemented in a child language learning study.

During this experiment, only a very limited amount of input was presented to participants. For a trained population like our adult group, which contained a large number of students of linguistics, this input was sufficient to pick up the correspondence rule. For more naïve populations, like children, this input clearly did not suffice. Because the goal of this experiment was to investigate the workings of the reward system, the results on learning the particular correspondence rule are not of much importance for now. However, to investigate whether children have the cognitive ability to develop any meta-linguistic awareness during language acquisition, it is essential that we train children on a rule that they can learn. Also, it would be preferable if participants in an opt out experiment first undergo a training phase during which they do not have to perform any actions. Such a training phase would enable children to acquire a linguistic regularity properly, before they are actually tested on their awareness of what they have acquired. This could be achieved by developing a miniature language that has to be learned during a training phase (see de Graaff, 1997, Lichtman, 2016 or Andringa & Curcic, 2015 for examples of such a language), such as the language we presented in Chapter 2.

When developing a miniature language for the opt out paradigm, there are some limitations with regard to the possible rule in that language that has to be learned. The paradigm requires a rule that allows for two types of linguistic elements: one linguistic element that predicts what comes next and another linguistic element that does not come with such a prediction. Participants can only show strategic behavior if one linguistic element predicts what is coming next and it is appealing to opt in, while another element does not predict what comes next, giving the opportunity to opt out. Candidates for such a rule might be differential object marking (Aissen, 2003) or certain classifier systems (Passer, 2016), like the one we described in Chapter 2.

As mentioned before, when using the opt out paradigm, it is extremely important to have a well-balanced reward system that rewards opting out and opting in differently, especially because children react differently to awards in comparison to adults (Frederick, 2005; Eigsti et al., 2006; Mischel et al., 1989). If the reward a participant can earn for making a correct decision is insufficiently appealing, it will be attractive to opt out regardless of awareness of the target structure. Conversely, when the reward for opting out is insufficiently appealing, participants might be keen to take a risk regardless of their awareness of the linguistic regularity that was present in the input. Because all children both opted in and opted out in our experiment, our current rewarding procedure meets our desired criteria, and should thus be maintained as much as possible in future adaptations.

The results from this chapter are promising, and show that the opt out experiment could be a valuable new tool to investigate meta-linguistic awareness. If the rule, training and rewarding system are controlled for carefully, this method could overcome some of the problems that direct subjective methods of investigating awareness have. Participants do not have to reflect on or verbalize their knowledge of the target structure. Therefore, this method can potentially probe into A-awareness, and provide new opportunities to make claims about the possible role of awareness in child language acquisition. In the next chapter we thus show how this method can be combined with the miniature language from Chapter 2 to investigate whether children develop A-awareness when acquiring grammatical elements.

Chapter 4

Do kindergarteners develop awareness of statistical regularities?[*]

We have seen in Chapters 1 and 2 that many studies suggest that detecting statistical regularities in linguistic input plays a key role in language acquisition (e.g. Erickson & Thiessen, 2015; Frost & Monaghan, 2016; Romberg & Saffran, 2010). Cumulating evidence suggests that learners' ability to pick up statistical regularities within linguistic input enables them to detect words within a continuous stream of speech, to map meanings onto these words, to group these words into abstract categories, and to determine the grammatical relations between these categories. Evidence that such statistical learning plays a role in different domains of language comes from several types of results. Artificial grammar learning experiments have shown that tracking statistical regularities allows learners to infer word boundaries and establish dependencies between linguistic elements (e.g., Aslin et al., 1998; Gómez & Gerken, 1999; Saffran et al., 1996 for infants; Gómez, 2002 for infants and adults; Endress & Bonatti, 2007; Perruchet & Vinter, 1998; Thiessen et al., 2013 for adults). Cross-situational word-referent learning tasks indicate that statistical learning plays a role in mapping form onto meaning (e.g., Frank et al., 2009; Frank et al., 2013; Smith et al., 2014; Smith & Yu, 2008; Vlach & Johnson, 2013 for infants; Kachergis et al., 2014; Vouloumanos, 2008; Yu & Smith, 2007 for adults). Several studies also show that distributional properties in the input can help learners to determine

[*] This chapter is a slightly modified version of an accepted article: published article: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (2021). Do kindergarteners develop awareness of the statistical regularities they acquire? *Language Learning*, *71(2), 573-611*.

the semantic category of words (e.g. Lany, 2014; Lany & Saffran, 2013 for infants; Chen et al., 2017; Monaghan et al., 2015 for adults).

In Chapter 3, we already discussed that although statistical learning is not implicit by nature (Batterink et al., 2015; Bertels et al., 2015), many scholars have defined it as learning that happens without awareness (e.g., Arciuli & Simpson, 2012; Aslin et al., 1998; Baker et al., 2004; Kidd, 2011; Reber, 1967; Rebuschat, 2013). Rebuschat and Williams (2012) note that, from time to time, statistical learning is explicitly equated with implicit learning (Conway & Christiansen, 2006; Perruchet & Pacton, 2006; Turk-Browne et al., 2005) and that the two are sometimes even concatenated, when speaking of 'implicit statistical learning' (e.g., Conway et al., 2010; Kidd, 2011). As statistical learning likely plays an important role in child language acquisition (e.g., Erickson & Thiessen, 2015; Frost & Monaghan, 2016; Romberg & Saffran, 2010), it is often assumed children acquire a target language without any awareness of the regularities in it. This might be reasonable to assume, because children are probably unable to verbalize or explicitly reflect on their knowledge. Many scholars of different theoretical persuasion indeed make this assumption (Aslin et al., 1998; Chomsky, 1986; Radford, 2004; Saffran et al., 1996; Tomassello, 2003; Ullman, 2016; Wijnen, 2013).

The results presented in Chapter 2 are compatible with the idea that statistical learning happens implicitly. In the experiment described in that chapter, kindergarteners had to acquire a meaningful grammatical element without explicit instruction. In the exit interview after the test phase of this miniature language learning experiment, no child could verbalize their knowledge of the grammatical element they had learned. This finding could indicate they indeed acquired the marker implicitly. However, in Chapter 3, we also argued that children might be aware of the knowledge they acquire even though they might not be able to verbalize it, and that research so far has not tapped into such awareness.

This chapter follows up on this discussion, and investigates whether statistical learning in young children is indeed only implicit, as often assumed. We trained kindergarteners on the same miniature language as in Chapter 2, and adapted the opt out task from Chapter 3 to measure whether kindergarteners are aware of their knowledge of an acquired meaningful grammatical marker. In the opt out paradigm, participants demonstrate such awareness by expressing uncertainty through a non-verbal response: by opting out. The results presented in this chapter are compatible with those from Chapter 2 suggesting kindergarteners can acquire this marker from distributional properties in the input. Furthermore, although none of the children could verbalize knowledge of the structure during exit interviews, their behavior during the opt out task indicates that they did develop awareness of it. These results thus nuance the assumption that statistical learning is an implicit process that happens without any awareness involved. Yet, before we can establish this, we need to elaborate on why it is necessary to investigate such issues in the next two sections.

4.1 Awareness and language acquisition

Awareness is studied within many disciplines of cognitive science, such as visual perception (e.g. Lamme, 2010), artificial intelligence (e.g. Dennett, 1993) and animal cognition (e.g. Proust, 2013). Informally, awareness can be described as the mental state in which living creatures know what they are experiencing. This conceptual representation of awareness stems from the classical Cartesian view on awareness (Descartes, 1637/1968), which assumed humans are aware of experiences if they can remember them coherently and can verbalize these experiences. This view has been labeled the Cartesian Theater (Dennett, 1993). In this fictive theater, living beings are spectators who are only aware of their experiences once they appear on the stage. If an experience does not appear on stage, the spectator is not aware of it at all. Hence, awareness in this classical view is an all-or-nothing phenomenon; someone is either aware of something or not.

However, readers may recall from Chapter 3 that alternative models of awareness have been proposed that question the importance of verbalization, coherence and memory as necessary requirements for awareness (Allport, 1988; Dehaene, Lau & Kouider. 2017; Dennett, 1993; James, 1890). This work assumes that awareness probably exists in different gradations, unlike the dichotomous characterization in the Cartesian Theater (Cleeremans, 2008, 2011, 2014). Cleeremans (2011), for example, distinguished between awareness at a phenomenal level (P-awareness) and awareness at the level of access (A-awareness). People possess P-awareness when they have a subjective experience of the object of awareness. Applied to linguistics, this means they are phenomenally aware when they are able to explicitly describe certain linguistic patterns: meta-linguistic P-awareness, which bears resemblance to what has often been called explicit knowledge in the field of second language acquisition (Hulstijn, 2005; Rebuschat, 2013). People possess A-awareness when they have an experience that is accessible to their cognitive system, but cannot be expressed subjectively. People have meta-linguistic awareness at access level if they perceive a particular linguistic pattern is present, and can (verbally) detect that 'something' is going on, without being able to determine which pattern this is exactly. This A-awareness resembles what has been called *noticing* in the L2 literature (Schmidt, 1990) and could be described as a mental state that floats somewhere between explicit and implicit knowledge.

Within linguistics, awareness has been studied most extensively within SLA, often focusing on the distinction between implicit and explicit knowledge (e.g. Andringa & Rebuschat, 2015; Hulstijn, 2015; Suzuki & DeKeyser, 2017). Explicit knowledge can be described as knowledge that is available to awareness, whereas implicit knowledge is not available to our awareness (Rebuschat, 2013). In Chapters 1 and 3, we have seen that research and findings on this topic are diverse (e.g. DeKeyser, 2003; Ellis, 2015; Han & Finneran, 2014; Paradis, 2009). As a result, several theories have been formulated about the role awareness could play when acquiring the grammatical structures of a second language: some authors claim that awareness of such structures is necessary to acquire them (DeKeyser, 2003; Schmidt, 1990), others suggest awareness merely facilitates the acquisition process (Ellis, 2003), and again others have argued awareness has no influence on the implicit linguistic knowledge that is gained (Krashen, 1981). The role of awareness in early language acquisition has been studied too, but perhaps not as extensively as in SLA. Many studies have investigated the role of awareness and explicit knowledge in the acquisition of vocabulary knowledge (e.g. Coyne et al., 2009; Silverman, 2007; Vaahtoranta et al., 2018), with a meta-analysis showing that vocabulary interventions in kindergarteners are effective and that more explicit instruction conditions result in bigger effect sizes (Marulis & Neuman, 2010). Furthermore, phonological awareness is widely studied in young children (e.g. Degé & Schwarzer, 2011; Gillon, 2018), often stressing its important relation to reading (e.g. Furnes & Samuelsson, 2011; Johnson & Goswami, 2010; Saygin et al. 2013). Nevertheless, the general theoretical debate about the role of awareness during language learning that is present in SLA seems absent in the field of child language acquisition in general, and statistical learning in particular.

We think there are good reasons to pursue the debate about the role of awareness in the fields of child language acquisition and statistical learning as well. Although the role of awareness is often not addressed explicitly in statistical learning studies, results from a study by Kerz, Wiechmann and Riedel (2017) suggest this topic requires more thorough investigation. In an artificial language experiment, they studied the acquisition of a grammatical marker and found that adults who did not possess any awareness of what they were learning scored slightly above chance (53-56%) in a two-alternative forced choice task measuring their acquired knowledge, but that they almost reached ceiling scores (88-91%) when they did develop such awareness. These results suggest that awareness plays a pivotal role in the L2 acquisition process in adults. The possibility that awareness plays a similar role when young children acquire grammatical elements is an open question that merits investigation. Children might develop some degree of awareness of the regularities they acquire, and this awareness might facilitate their acquisition process. They could also develop awareness, but not use such awareness in their learning process. Alternatively, awareness could be a prerequisite for learners to acquire a language in the first place. These questions are rarely asked, because the common assumption seems to be that when children engage in statistical learning, this happens implicitly (e.g. Arciuli & Simpson, 2012; Aslin et al., 1998; Baker et al., 2004; Conway et al., 2010; Erickson & Thiessen, 2015; Frost & Monaghan, 2016; Kidd, 2011; Perruchet & Pacton, 2006; Reber, 1967; Rebuschat, 2013; Romberg & Saffran, 2010; Wijnen, 2013). The assumption that children do not develop awareness when learning from statistics seems mainly held because they are unable to verbalize their awareness and seemingly do not possess meta-linguistic P-awareness. Yet, in light of the graded nature of awareness assumed in this book (Cleeremans, 2008, 2011, 2014; Dehaene et al., 2017; Dennett, 1993), it is conceivable that children develop meta-linguistic awareness at a non-verbal level. To our knowledge there is no research investigating whether this is the case. To gain insight into this issue, we would need to tap into meta-linguistic A-awareness.

4.2 Measuring awareness

Although measuring awareness is a difficult enterprise, research on adults provides us with several methodological tools to do so (see Rebuschat, 2013 for a review). Awareness can be measured using retrospective verbal reports in which researchers ask learners to verbally reflect on what they are learning (e.g. Dienes, Broadbent & Berry, 1991; Rebuschat & Williams 2012; Williams, 2005). If learners can verbalize their linguistic knowledge in such reports, this is taken to mean that they are aware of it. Alternatively, subjective measures, such as confidence judgement tasks, can be used where learners have to indicate how confident they are of their task responses (e.g. Dienes, Altmann, Kwan & Goode, 1995; Hamrick & Rebuschat, 2012). If learners are confident of their task behavior, this is taken as a sign that they are aware of the knowledge underlying this behavior. Another possible way to probe into awareness is the contrastive use of direct and indirect tests (e.g. Jiménez, Méndez & Cleeremans, 1996; Reed & Johnson, 1994). The former type of tests explicitly asks learners to rely on knowledge they are aware of when executing a task, whereas the latter does not. If learners demonstrate they possess knowledge on one of the two tasks, but not on the other, this contrastive behavior can be used to make inference about the status of the knowledge. Timmermans and Cleeremans (2015) label these types of measures 'direct subjective methods', as they directly assess the awareness a person has, by relying on their subjective experience. As such, these measures of awareness mainly tap into P-awareness: the type of awareness that people have when they possess a coherent experience of their object of awareness and can verbally reflect on it.

However, direct objective methods have limitations (Rebuschat, 2013). Retrospective verbal reports, in which participants verbally reflect on whether or not they were aware of the gained knowledge, tend to overlook the fact that subjects can develop such awareness without being able to translate their knowledge into words (Dienes & Berry, 1997; Dienes & Fahey, 1995). Subjective measures are prone to a response bias, as participants set their own criteria for responding (Dienes, 2004, 2008; Reingold & Toth, 1996). Furthermore, participants' behavior on direct tests might be confounded by unconscious knowledge (Merikle, Smilek & Eastwood, 2001; Reingold & Merikle, 1988, 1990). Because of these limitations and because we cannot expect children to verbalize rules or partake in confidence judgement tasks, as they likely do not possess P-awareness, it seems unreasonable to employ these methods to test awareness in children. Other lines of research into awareness in young children, such as awareness of language itself at a macro-level (e.g., Atagi & Sandhofer, 2020) or metalinguistic awareness about written language (Ke, Miller, Zhang & Koda, 2020) are also limited in that they do not give us methods for probing into awareness of grammar during exposure to a new spoken language. Several tasks such as grammaticality judgment tasks or wug-tests, have been argued to measure meta-linguistic awareness and might overcome these issues (Barac & Bialystok, 2012; Bialystok, 1986). Nevertheless, although such methods potentially require participants to possess meta-linguistic A-awareness, we do not necessarily know whether participants are aware of *what* the regularities are that they have acquired, only that there *are* regularities. Behavior on such tasks could equally well be the result of completely implicit knowledge of the rules. Therefore, if children possess awareness at the A-level, the question is what type of measurement we need to establish this, when we cannot use direct subjective methods, grammaticality judgement tasks or wug-tests.

One possibility for establishing whether awareness of the target regularities is involved when young learners acquire a language, would be the use of indirect objective methods. These are methods that do not require subjective introspection by the participants (Timmermans & Cleeremans, 2015). An example of an indirect objective method that can be used to measure A-awareness is the opt out paradigm, which we introduced in Chapter 3. In this paradigm, participants can demonstrate awareness by expressing uncertainty through a non-verbal response: the so-called uncertainty response. Typically, participants in such a task perform a cognitive task with easy and difficult trials. If they possess awareness of their cognitive capacities, they should express their uncertainty for the more difficult trials by choosing the uncertainty response. In this paradigm, participants obtain a reward for a correct decision in a trial and are able to express their uncertainty by opting out of a trial. In the latter case, participants will always receive a reward, but this reward is smaller than the one they get for a correct decision. If participants are aware of their cognitive capacities, they should take the risk to obtain a bigger reward in easy trials and opt out and obtain a smaller, but certain, reward in the more difficult trials.

In Chapter 3, we also discussed several experiments that used this paradigm to investigate awareness in a variety of non-human animals (e.g. Hampton, 2009; Smith, 2009; de Waal, 2016). These experiments have shown that rhesus monkeys are aware of their memory of particular stimuli (e.g. Hampton 2001), as are macaques (e.g. Fujita, 2009) and orangutans (e.g. Suda-King, 2008). Furthermore, rats are able to judge how well they distinguish long from short lasting noises (e.g. Foote & Crystal 2007), pigeons can indicate how good they are at classifying sparse and dense pictures (e.g. Sole, Shettleworth & Bennett, 2003) and dolphins can assess their own performance when discriminating between high and low pitched sounds (e.g. Smith, 2010). Importantly, none of these animals had to engage in subjective introspection with regard to their cognitive capacity, but instead performed objective observable behavior that was the indirect effect of their awareness. Thus, the opt out paradigm seems a useful indirect, objective measure to tap into A-awareness in children.

4.3 The current study

In this study, we exploited the opt out paradigm to measure meta-linguistic awareness in kindergarteners. This work extended the studies that have been discussed in Chapters 2 and 3. Chapter 3 investigated the balance in rewards for correct decisions and opt out responses, to avoid that the uncertainty response would be too appealing and bias participants to pick the reward regardless of their awareness. Several studies have shown that children are much more likely to go for a reward that they are guaranteed to get rather than taking an option for an uncertain reward, because they cannot delay their gratification in an adult-like manner (Eigsti et al., 2006; Mischel et al., 1989). The experiment from chapter 3 showed that the rewards used in the present study strike the right balance, as the children who participated did not show a bias towards the uncertainty response.

In Chapter 2, we presented an artificial language learning experiment in which the reward system could be incorporated. In this artificial language kindergarteners learned a grammatical marker, which could serve as a target structure that children could potentially develop awareness of. This grammatical marker expressed number when it was combined with a noun and its function could be learned from distributional properties in the input. Children seemed able to acquire this marker using statistical regularities and post-test verbal reports did not indicate the children were aware at the phenomenal level (P-awareness) of their acquired knowledge.

In this chapter, we present a study that combined the artificial language learning task and the opt out procedure. The main goal of this study was to see whether children develop A-awareness when they acquire a grammatical pattern on the basis of statistical regularities in the input. To test this, we thus used a novel measure in which children can demonstrate such awareness by expressing uncertainty through a non-verbal response. If children develop awareness of their acquired knowledge, they should opt out more often on trials that are difficult and they should opt in on trials that are easy, when they have acquired this marker.

In the spirit of recent calls for more replication (Marsden et al., 2018), and keeping in mind that a single study can never be conclusive and accidental findings might be published (Cumming, 2014), a secondary goal of our study was to replicate the results from Chapter 2 that young learners are indeed able to acquire a meaningful grammatical marker from distributional properties in the input. Replicating our earlier findings would provide more evidence for the idea that children can use distributional cues to learn grammatical regularities (e.g. Lany, 2014; Lany & Saffran, 2013).

4.4 Method

4.4.1 Participants

70 native Dutch speaking children (35 males, 35 females, *M* = 5;5 years, *SD* = 0;6, *range* = [4;3 - 6;10]) took part in this experiment. All children were in kindergarten and were recruited from three primary schools in the Central and Western area of the Netherlands. They did not have any diagnosed language or communication disorders. No further information about the children's linguistic background was collected. Seven participants had to be excluded, because they provided answers before they were exposed to the test items, probably due to a lack of concentration. These children consistently pointed towards the part of the screen they thought was showing the correct picture, before the pictures were on the screen. As a result, we present data from 63 children (32 males, 31 females, *M* = 5;6 years, *SD* = 0;5, *range* = [4;3 - 6;10]).

4.4.2 Materials

4.4.2.1 Miniature language and target structure

We used the artificial language already described in Chapter 2, which consisted of four proper names, three verbs, two grammatical markers, six frequent nouns, twelve infrequent nouns and one conjunction. Apart from the proper names, which might occur in Dutch, all words were novel words. The language had subject-verb-object word order. In this language, a noun phrase on its own does not encode number and could correspond to both singular and plural referents. In a sentence, however, an argument noun phrase must be introduced by the nominal marker *tra*. This type of nominal marker included in our artificial language is quite common in languages with residual noun classes, such as the Kwa-language Akan (Appah, 2003) or the Austronesian language Cebuano (Parnes, 2011). In our miniature language, a noun introduced by *tra* can refer to both singular and plural referents: the correct interpretation must be inferred from the visual context. The sole function of *tra* therefore is to turn a bare noun into an argument. The language includes another nominal grammatical marker *pli*, which encodes number and indicates that the noun necessarily refers to multiple referents. A similar type of marker appears in the Akan-language Gungbe (Aboh, 2004; Aboh & DeGraff, 2014). The marker *tra* can be seen as a default nominal marker with no number specification, whereas *pli* not only turns the noun into an argument but adds information about number. In short, the rule participants had to learn was that whenever *pli* preceded a noun, this noun always referred to multiple referents

4.4.2.2 Structure of the experiment

The overall structure of the experiment was almost completely similar to the experiment described in Chapter 2. To learn the regularity, participants were told a story with four protagonists (Carlo, Julia, Marco and Maria), who were going on a holiday to a country of which they did not speak the language. Participants were asked to help the protagonists learn this new language. They were told they would see pictures and hear things in the new language that matched the pictures they saw. The experiment consisted of four parts. It started out with a small vocabulary training. After the vocabulary training a rule training session followed. During this session, children received input to acquire the grammatical rule. After the rule training, participants performed a picture matching task to test whether they acquired the rule. In the picture matching task, participants heard sentences from the rule training phase, and had to choose which of two pictures matched each sentence. For the test items of this task, participants had to choose between pictures with either one or multiple referents. The two pictures always referred to different referents. 12 of the experimental items contained *pli*. For these items, the target picture always showed multiple referents. The other 12 of the experimental items contained *tra*. Any number of referents were grammatical for these items. Sensitivity to the statistical regularity in the input would lead to more target answers on trials with *pli* than on trials with *tra*, because for trials with *pli* participants could base their response on the number of referents the picture showed, whereas the number of referents was not indicative for trials with *tra*. After the picture matching task, participants continued with the opt out phase to test whether they were aware of the rule. Since the majority of this experiment is similar to that in Chapter 2, we refer the reader to that chapter for a full description of the experiment. Apart from the following three points, the present and previous experiment were exactly the same, and were conducted by the same experimenter. An overview of the experiment can be seen in Figure 4.1.

1. The vocabulary test after the vocabulary training at the start of the present experiment contained only two test items per noun. In our previous experiment, participants continued to receive test items until they identified all six nouns correctly four times. This proved frustrating, however, because most children had difficulties learning all nouns from the limited input they received. Therefore, it was problematic to keep the vocabulary training as it was. The purpose of the vocabulary test was to familiarize participants with the picture matching task, and participants did not receive any input to the grammatical regularity at this point. We think it is unlikely that this change would have affected learning of the grammatical regularity.

2. All participants in the present study performed the attention task, whereas only half of the participants performed this task in our previous study. In our previous experiment, we observed a tendency towards better performance by children in the attention task condition. Therefore, we decided to include the attention task for all children.
3. The opt out phase included in this study, in which we assessed whether children developed awareness of the grammatical regularity, was included in this experiment. It came after the picture matching task and cannot have affected performance on this task. The opt out phase took place before the informal debriefing, and could have influenced the answers children gave during this debriefing. This opt out phase is described in the next section.

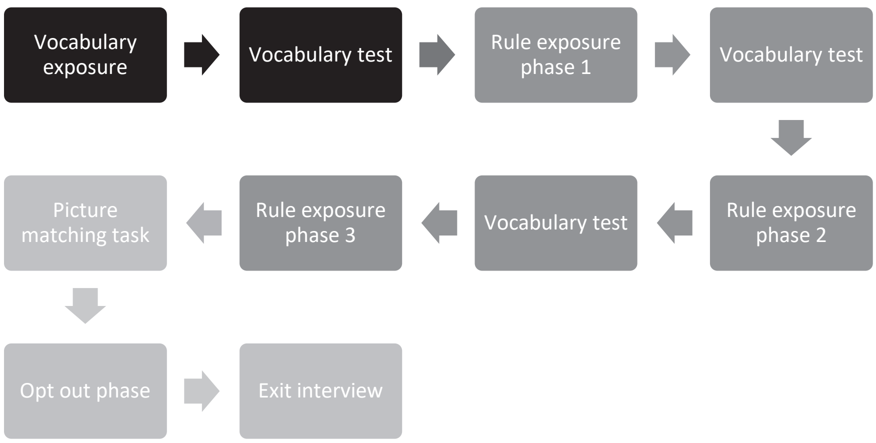


Figure 4.1. A visual representation of the structure of the experiment. The darkest boxes show the parts of the vocabulary training. The lighter boxes show the different parts of the rule training. The lightest boxes show the test phase.

4.4.2.3 Testing awareness of the rule

After the picture matching task, the opt out phase started in order to determine whether participants were aware of the grammatical pattern they had acquired. Before this phase started, participants engaged in a small practice session. This practice session was necessary to prepare the children for the opt out phase of the experiment and familiarize them with the rewards they could obtain in that phase. In this session, participants were exposed to six frequent nouns and had to choose between two pictures. Importantly, for every correct decision a child would get a reward. Children helped the protagonists cross a river with 10 stepping stones to return home after their holiday. The protagonist would move 2 stones for every correct answer. Children received a sticker for every protagonist that returned home in this practice session and in the opt out phase. In the opt out phase, participants heard the same test items that were used in the picture matching task, up to and including the grammatical marker, but without the last word, while seeing a blank screen. After hearing these incomplete sentences, participants were given two options. They could hear the whole sentence including the last word. When hearing the full sentence, they would see the two pictures and they would have to make a decision between these pictures. As before, if they decided correctly they would earn 2 steps and if they made a wrong decision, they would not obtain any reward. Participants could also choose to opt out. In this case, they decided not to hear the full sentence, but simply to move on to the next test item. When opting out, the protagonist would move 1 stepping stone. Thus, in order to bring home as many protagonists as possible, participants had to consider their chances of providing a correct answer. If participants decided to hear the full sentence, the sentence would always end with one of the infrequent nouns that occurred in the training phase. This opt out phase consisted of the same 24 test items from the picture matching task, but with the nouns removed. Thus, 12 items ended after *pli*, and 12 items ended after *tra*. When participants decided to hear the full sentence, they would hear a full sentence. If a participant decided to hear a full sentence on each occasion, this participant would hear all 24 test items from the picture matching task.

After the opt out phase, a debriefing took place in which we examined whether participants were aware of their knowledge in a way they could verbalize. During this debriefing, participants were asked how they knew what the correct answer during the picture matching task was, whether they knew the meaning of the words *pli* and *tra* and why they decided to opt in or opt out.

4.4.3 Procedure

All stimuli were recorded by a female native speaker of Dutch. The test was administered in a quiet room at the participants' school. The task was presented on a laptop using E-prime (Psychology Software Tools, 2012). During vocabulary training, nouns and their accompanying pictures were presented for three seconds before automatically moving to the next noun. As the goals of this training were familiarization and keeping attention, we did not register scores during the vocabulary test systematically enough to report on. During rule training, sentences and their accompanying pictures were presented for four seconds before automatically moving to the next sentence. During both test phases, the experimenter pressed a button on the keyboard that corresponded to the answer the participant gave. Scores from this phase were registered automatically. Both the participant and the experimenter listened to the audio using headphones. The vocabulary training lasted 8 minutes on average, the rule training phase 15 minutes, the picture matching task 5 minutes, and the opt out phase 5 minutes. The full experiment took approximately 30 minutes per participant. Ethical approval for this study was obtained from the University of Amsterdam and passive consent was obtained from children's parents or legal guardians before the start of the study.

4.4.4 Analysis

All analyses were carried out in R (R Core Team, 2015) using the lme4 package (Bates et al., 2015) where needed. To determine whether participants grasped the target regularity and whether they were aware of this regularity, a logistic regression model with mixed effects was carried out for the picture matching task and opt out phase separately. We expected that if children learned the regularity, they should score better on sentences with *pli* than on sentences with *tra* in the picture matching task, and if they were aware of this regularity, they should opt out more often when sentences contained *tra* than when sentences contained *pli*. Both models took the responses from the tasks (in the picture-matching task, 1 for a correct answer, 0 for an incorrect answer; in the opt-out phase, 1 for opting out, 0 for not opting-out) as a dependent variable, marker type as a within-participants fixed effect, participant as a between-participants random effect and test item as a within-participants random effect. Our fixed effects were included in these models, because we were *a priori* interested in their contribution to the outcome (Gelman & Hill, 2007). In both models, orthogonal sum-to-zero contrast coding was applied to our binary fixed effect (i.e., marker type; Baguley, 2012, p590-621). As we aimed to keep the models as fully specified as possible by including random intercepts for participants and test items, (Barr et al., 2013), we had to increase the number of possible iterations to 100.000 (Powell, 2009) to solve issues with non-converging models. This enabled us to report on design driven random effect structures that were still justified by our data (Jaeger, 2009). In line with Baguley (2009), we report simple rather than standardized effect sizes and confidence intervals. As an exploratory analysis, we conducted a Pearson correlation to investigate possible relationships between the scores on both parts of this experiment.

4.5 Results

The descriptive statistics from the picture matching task and opt out phase can be found in Figure 4.2 and Table 4.1. We first tested our expectation that children were sensitive to the distributional properties in the input. If children were sensitive to these properties, they should score better on sentences with *pli* than on sentences with *tra* in the picture matching task. Results from this task showed that participants gave more target answers when sentences contained *pli* than when sentences contained *tra* (OR = 1.344, 95% CI = [0.980, 1.843], z = 1.834, p = .067), but this effect was not significant. Based on this sample, we cannot claim that kindergarteners became sensitive to this regularity, although we see a tendency towards such an effect.

In the opt-out phase, we tested whether children developed A-awareness of the target structure. If they developed such awareness, they should opt out more often when sentences contained *tra* than when sentences contained *pli*. Results from the opt out phase showed participants did so; they opted out more often when sentences contained *tra* than when sentences contained *pli* (OR = 1.482, 95% CI = [1.089, 2.017], z = 2.502, p = .012). This suggests that children were at some level aware of this regularity. Note that the ranges we observed in opt out scores vary greatly. This seems to indicate that, irrespective of whether these children favor to reject the uncertainty response or not, thus being risk takers or not, they were aware of the regularity. Some children were likely to opt out in general, whereas other children took a risk more often and hardly opted out, but as the type of marker was a within participant factor, such personal preferences did not influence this main effect.

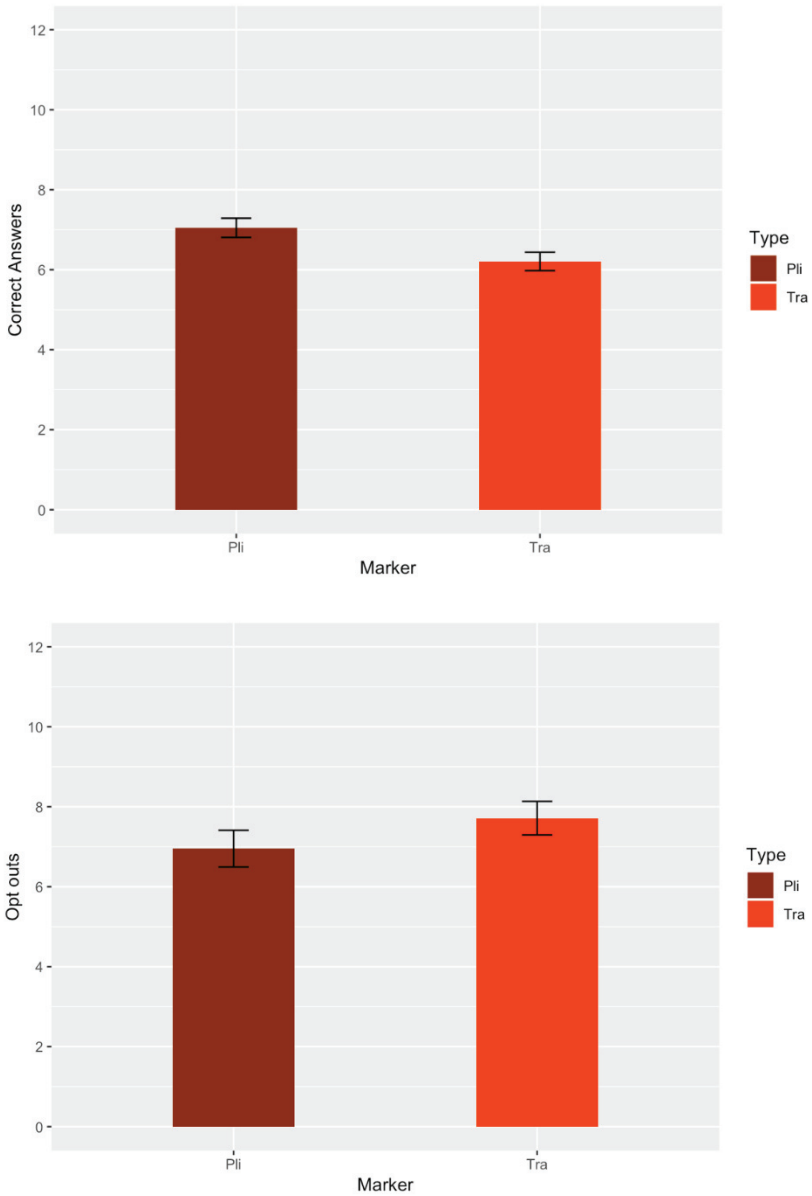


Figure 4.2. Graphs depicting the results from the picture matching task (above) and the opt out phase (below). Scores from the picture matching task indicated the number of target answers produced and scores from the opt out phase indicated the number of times participants opted out. Scores could range from 0-12.

Table 4.1. Scores from the picture matching task and the opt out phase.

	Picture matching task			Opt out phase		
	M	**SD**	**Range**	**M**	**SD**	**Range**
Pli	7.05	1.90	3-12	6.95	3.68	0-12
Tra	6.21	1.82	2-10	7.71	3.37	0-12

Note: Scores from the picture matching task indicate the number of target answers produced and scores from the opt out phase indicate the number of times participants opted out. Scores could range from 0-12.

In the debriefing, when asked how they reached a decision on the picture matching task, 75% of the participants (*N* = 47) reported they did not know how they made a decision. The other 25% of them (*N* = 16) claimed that they had heard the sentences before and remembered what they meant. When asked for the meaning of *pli* and *tra*, 71% of the participants (*N* = 45) said they did not know what these words meant, another 21% of them (*N* = 13) gave the meaning of a noun they learned during the exposure, and finally, 8% of the participants (*N* = 5) came up with some other meaning for these words. This suggests that children were not phenomenally aware of the regularity they acquired. When asked why they opted out or opted in, 22% of the children (*N* = 14) said they based their decision on the difficulty of the trials, but none of them linked this to plurality. The other 78% of the children (*N* = 49) did not report that they based their opt out behavior on the difficulty of the trials.

To determine whether a correlation existed between awareness in the opt out phase and learning in the picture matching task, a difference score was calculated for both phases. For the picture matching task, this score was calculated by subtracting the number of target answers on sentences with *tra* from the number of target answers on sentences with *pli*. A larger difference in a positive direction indicates that the participant gave more target answers on sentences with *pli* than on sentences with *tra*. Difference scores per participant for the picture matching task can be seen in Figure 2. For the picture matching task, 54% of the participants (N = 34) exhibited a positive difference score, 16% of the participants (N =10) a neutral difference score, and 30% of the participants (N = 19) a negative difference score. For the opt out phase, this score was calculated by subtracting the number of times a participant decided to continue when a sentence ended with *tra* from the number of times a participant decided to continue when a sentence ended with *pli*. If this difference score was higher in a positive direction, this indicates the participant opted out more often when a sentence ended with *tra* than when a sentence ended with *pli*. Difference scores per participant for the opt out phase can be seen in Figure 4.3 as well. In this phase, 49% of the participants (N = 31) exhibited a positive difference score, 32% of the participants (N = 20) a neutral difference score, and 19% of the participants (N = 12) a negative difference score. A Pearson's correlation test did not show a significant correlation between learning and awareness in this experiment (r = -0.133, p = 0.23).

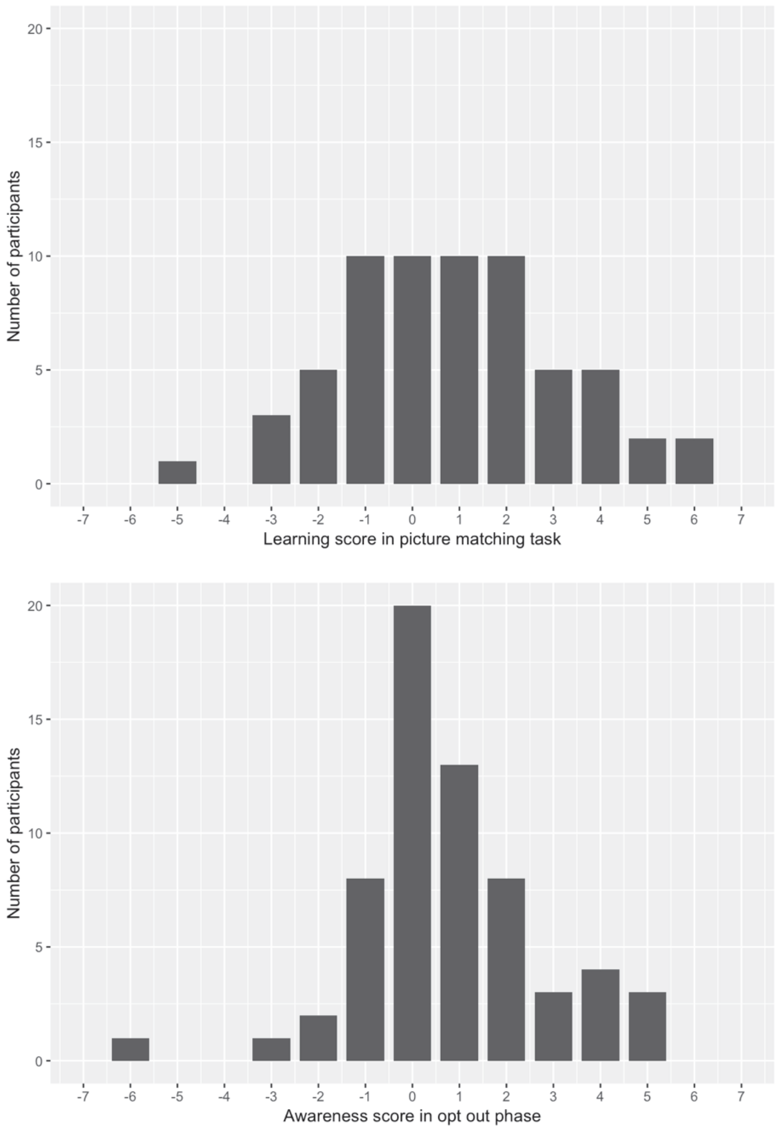


Figure 4.3. Graphs depicting the difference scores from the picture matching task (top) and the opt out phase (bottom) per participant. A positive score in the picture matching task indicates a participant gave more correct answers for sentences with *pli* than sentences with *tra,* and thus indicates learning of the regularity. A positive score in the opt out phase indicates a participant opted out more often for sentences with *tra* than sentences with *pli*, and thus indicates awareness of the regularity.

4.6 Discussion

The main goal of this opt out experiment was to see whether kindergarteners develop awareness of the linguistic regularities they acquire using distributional properties of the input. We wanted to investigate this since statistical learning is not necessarily implicit in nature, though many scholars assume it is (e.g. Arciuli & Simpson, 2012; Aslin et al., 1998; Baker et al., 2004; Conway et al., 2010; Erickson & Thiessen, 2015; Frost & Monaghan, 2016; Kidd, 2011; Perruchet & Pacton, 2006; Reber, 1967; Rebuschat, 2013; Romberg & Saffran, 2010; Wijnen, 2013). To our knowledge, there have not been systematic investigations into the implicit or explicit nature of statistical learning with children. While investigating whether young children develop awareness during learning, this study also provided an opportunity to try to replicate our findings from the study presented in Chapter 2, in which we demonstrated that a number marker could be learned on the basis of statistical information.

In this study, we showed that participants gave more target answers when sentences contained *pli* than when sentences contained *tra*, but that this effect was not significant. This suggests we were unable to replicate our earlier finding that kindergarteners use distributional properties from the input to acquire a number marker. However, replication is not necessarily a matter of significance, and findings might be replicated in several ways. Marsden et al. (2018) explain that replication can be achieved by carrying out narrative comparisons, comparing descriptive statistics, or interpreting dichotomous findings from null hypothesis significance testing. Reasonable arguments might be given for each type of comparison, and it is unclear which comparison is most appropriate. If we look beyond *p*-values and include effect sizes and confidence intervals in the comparison of this study and the previous, then the outcomes are not so different. In Chapter 2, we reported a main effect of grammatical marker in the same direction and found evidence that children could acquire this regularity (*OR* = 1.509, 95% CI = [1.040, 2.190], z = 2.167, p = .030). Perhaps it was more difficult to establish the effect in the current study with a particular degree of certainty, as it is slightly smaller than the effect in our previous study. Importantly, our current effect size (OR) falls within the 95% CI of our earlier study and vice versa: *OR* = 1.344, 95% CI [0.980, 1.843], z = 1.834, p = .067. Given each study, the results of the other can be expected as they fall within the range of possible results. Thus, this study can be seen as adding support to the view that kindergarteners use distributional properties of the input to acquire a meaningful grammatical marker. If these findings were to hold in new studies (see for example the following two chapters), this would indicate that statistical learning not only plays a role in word segmentation, word referent learning and the acquisition of grammatical markers (e.g. Frost & Monaghan, 2016; Romberg & Saffran, 2010), but is also involved in acquiring more grammatical structures in young learners (e.g. Lany, 2014; Lany & Saffran, 2013). However, caution is still needed in interpreting these results, because the p-values from both studies are 'dancing' around .05.

During the exit interviews none of the children could reflect verbally on the target structure and none could explain what either *pli* or *tra* meant. This seems to be in line with the idea that young children do not develop awareness of the regularities in the target language (Aslin et al., 1998; Chomsky, 1986; Radford, 2004; Saffran et al., 1996; Tomassello, 2003; Ullman, 2016; Wijnen, 2013). However, awareness likely exists at multiple levels, and not only in ways that can be verbalized (Allport, 1988; Cleeremans, 2008, 2011, 2014; Dehaene et al., 2017; Dennett, 1993; James, 1890). Perhaps children are aware of the regularity, but cannot verbalize this during an interview. To measure awareness at another level, we included the opt out procedure and found that children opted out more often when they heard a sentence containing *tra* than when they heard a sentence containing *pli*. This indicates that children were more readily inclined to make a decision on *pli* trials, for which they could be more certain. It also showed that they were less willing to try to obtain a reward for the *tra* trials, possibly because they were less certain about their ability to give a correct answer in these trials. This pattern of results is what one would expect if children were aware of the target structure. When the structure is acquired, trials with *pli* are predictable, and when a participant is aware of this regularity, it is expected that this person does not opt out in these cases, but keeps playing. Sentences with *tra* are unpredictable, even when the regularity is acquired, because *tra* can refer to both singular and plural. In these cases, children who are aware, are expected to opt out more often. We observed precisely this behavior, but as the effect is only small and borderline significant, we are cautious in making too strong claims on the basis of this result. Replicating this finding is highly necessary.

If our findings indeed indicate that children develop awareness of their acquired knowledge using statistical learning, this would have some important implications. Mainly, it would mean that statistical learning is not necessarily as implicit as it is often assumed. This result could be a first step to conduct more research about the role of awareness in early language acquisition, of the kind that is already taking place in the field of SLA (e.g., Andringa & Rebuschat, 2015; DeKeyser, 2003; Ellis, 2003; Hulstijn, 2015; Krashen, 1981; Schmidt, 1990). Perhaps young learners also need some form of awareness of such structures in a language to acquire them (DeKeyser, 2003; Schmidt, 1990), or it only helps their acquisition process, but is not necessary (Ellis, 2003). Of course it could also have no influence on the implicit linguistic knowledge that is gained at all (Krashen, 1981). Any of these scenarios could be possible for children as well, and a possible correlation between participants' learning and awareness might have shed light on this issue, but as we found a small negative correlation between these two it is difficult to make claims about the relationship between learning and awareness. The absence of such a correlation between learning in the first phase and opt out behavior in the second phase may seem counterintuitive, but this is not necessarily so. An absent correlation would be consistent with views that learning and awareness are uncorrelated (e.g. Krashen, 1981); a null effect like this, however, cannot be interpreted as such, and there may be other explanations for the absence of a significant correlation.

One explanation for the absence of a significant correlation between learning and awareness could be that the group we tested is non-ergodic (Lowie & Verspoor, 2019). This means not every individual is representative of the group and vice versa. Perhaps awareness plays some role in acquiring language in young learners, but the extent to which it does is subject to vast individual differences. This idea could be supported by the variation in difference scores from the picture matching task and opt out phase. Not every individual gave more correct answers for sentences containing *pli* than for sentences containing *tra* during the picture matching task and not every individual opted out more often for sentences ending with *tra* than for sentences ending with *pli* during the opt out phase. The answers participants gave during the informal debriefing also point towards individual differences in learning the markers. Some children did not know their meaning at all, whereas others thought they had a referential meaning.

The linguistic background of the children might play a role in explaining the individual differences in our study. Several authors have suggested, for example, that bilingualism might influence the development of meta-linguistic awareness (Barac & Bialystok, 2012; Bialystok, 1986), which might also be the case in our sample. In the same way as our study in Chapter 2, we also opted to not collect any data that did not pertain to our research questions in this study. We were only investigating whether children are able to develop awareness, and not necessarily whether all children develop such awareness or whether particular children develop more of this awareness than others. In this way, our sample was representative of the population we were interested in and served our specific goals (see Kruskal & Mosteller, 1979; Kukull & Ganguli, 2012 for more discussions about different types of representation), but does not enable us to draw firm conclusions about behavior in either specific individuals or specific populations. As a result, we were unable to explore whether the individual variation we observed in the present sample could be related to individual characteristics, such as the child's linguistic background. Hypotheses about possible relationships between awareness and participant features may be investigated in dedicated individual differences studies. Another reason we were unable to uncover a possible correlation is that the effect sizes for the picture matching task and opt out phase were small. Such small effect sizes in statistical learning studies are not uncommon (Kerz et al., 2017), and we may wonder whether increased awareness in this age group, like in adults, leads to enhanced learning scores. Future research, for example, could manipulate awareness in learners to see whether this leads to increased learning outcomes and distinct developmental patterns, which could even have impact in educational settings like kindergarten.

In all, this study was the first, to our knowledge, to apply indirect objective measures to investigate whether kindergarteners develop awareness of target structures during language learning. Such measures have great potential for measuring this as they do not require participants to verbalize their awareness, and potentially tap into different levels of awareness (e.g. Timmermans & Cleeremans, 2015). It would be worthwhile to develop these measures further to investigate this issue to a better extent. Hopefully, this study will lead to more research on the topic, such that clearer hypotheses can be formulated about the relation between awareness and language acquisition in young learners.

The goal of this chapter was twofold. First, we wanted to replicate the study presented in Chapter 2, in which we showed that kindergarteners might use statistical properties of the input to acquire a number marker. Our current results partially replicate these findings, as the current effect size is in line with our previous findings, but we cannot reject the null hypothesis based on our current sample. We argue that these two studies together are compatible with the idea that kindergarteners can acquire such a number marker on the basis of statistical learning processes. Second, we wanted to see how implicit the acquisition of such a marker actually is. Many scholars assume statistical learning is an implicit process that happens without any awareness involved. The opt out results seemingly nuance this idea. Although children were unable to verbalize their awareness of the grammatical regularities they acquire, they seemed to develop awareness of what these regularities are on some other level. What the exact role of this awareness in the acquisition process is, is still an open question. The studies presented in the following two chapters aim to provide more insight into the role awareness might play when children acquire grammatical markers.

Chapter 5

The effect of explicit instruction*

It is often noted there is a fundamental difference between young children learning the grammatical regularities of a new language and adults going through the same process (Bley-Vroman, 1990; Hartshorne et al., 2018; Krashen et al., 1982; Singleton & Ryan, 2004), as we mentioned briefly in Chapter 1. A central question in linguistics is how we can characterize this difference between adult and child language learners. Scholars have proposed that differences between child and adult language learning might be due to different linguistic inputs, contextual factors, general cognitive aptitude, or might arise from a critical period for language learning specifically (Andringa & Dabrowska, 2019; Birdsong, 2005, 2006; Birdsong & Van Hove, 2016; Brooks & Kempe, 2019; Doughty, 2019; Granena & Yilmaz, 2019; Hakuta, Bialystok & Wiley, 2003; Johnson & Newport, 1989; Newport, 2020; Singleton & Munoz, 2011). One take on this issue is that there is a difference in the available learning mechanisms for adult and child learners of language (Lichtman 2016; Pakulak & Neville, 2011; Zwart et al., 2017). In this scenario, children use more implicit or procedural learning mechanisms when they acquire grammatical structures, whereas adults rely on explicit or declarative learning mechanisms, and possibly in conjunction with implicit learning mechanisms (DeKeyser, 2000, 2003; R. Ellis, 2005, 2009; Paradis, 2004, 2009; Ullman, 2001).

* This chapter is a slightly modified version of a manuscript that is currently under review: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (2021). The effect of explicit instruction on implicit and explicit linguistic knowledge in kindergarteners. *Language Learning and Development*, https://doi.org/10.1080/15475441.2021.1941968.

Lichtman (2016) summarizes the idea that children and adults make use of different learning mechanisms in the form of the 'maturational hypothesis', whereby both children and adults are maturationally constrained to using particular learning mechanisms. She also observes that some scholars take a fairly radical position with regard to this hypothesis by claiming children learn grammatical rules using implicit learning mechanisms only (Bialystok, 1994; DeKeyser & Larson-Hall, 2005; Paradis, 2004, 2009). In Chapter 1, we already indicated that this is perhaps best illustrated by Bialystok (1994) who posits "rules make sense to adults; they make little difference to young children" (p.565), suggesting children do not rely on explicit learning when acquiring grammatical structures. However, very few studies have investigated specifically whether young children actually can learn grammar explicitly. It could well be the case that children utilize implicit learning only and do not rely on explicit learning mechanisms, but to our knowledge such claims have hardly been tested experimentally. In the present chapter we therefore investigate whether young children could benefit from awareness of form through explicit instruction of a grammatical regularity. We present an artificial language learning experiment in which we investigated if and how kindergarteners are affected by explicit instruction when they have to learn a meaningful grammatical element.

5.1 Explicit and implicit instruction, learning and knowledge

As we discussed in previous chapters, the difference between implicit and explicit language learning may be captured as follows. Explicit learning occurs when learners are aware that they are learning, and make an effort to discover the regularities in the target language, whereas implicit learning happens when learners do not know they are learning something and do not have such intentions (e.g. Dörnyei, 2009; N. Ellis, 2015; Hulstijn, 2015; Ullman 2001, 2016). Often, these types of learning are defined by the kinds of knowledge they result in (e.g. Hulstijn, 2015; Reber 1967), with explicit learning seen as learning that results in explicit knowledge, and implicit learning as learning that results in implicit knowledge. Explicit knowledge is available to awareness, whereas implicit knowledge is not (Rebuschat, 2013). Knowledge is regarded as explicit when learners can verbally reflect on it or when they are confident of their behavior during an explicit knowledge-based task. In contrast, learners might not be confident about their performance during a linguistic task, or they may be unable to verbally reflect on the knowledge they possess. In these cases, they are supposedly unaware of their linguistic knowledge, indicating it is implicit. Although explicit and implicit learning often result in explicit and implicit knowledge respectively (e.g. Andringa & Rebuschat, 2015; N. Ellis, 2005, 2011; Han & Finneran, 2014; Hulstijn, 2005, 2015), this need not be the case. Explicit learning might result in implicit knowledge, and implicit learning could lead to explicit knowledge (Batstone, 2002; DeKeyser, 2009, Williams, 2009). This occurs for example, when learners are aware that they are learning a grammatical structure in a second language, but nevertheless end up with knowledge they are not aware of (R. Ellis, 2009).

Furthermore, we can make a distinction between explicit and implicit instruction (DeKeyser 1995; R. Ellis, 2005, 2009; Housen & Pierrard, 2006). When receiving explicit instruction, learners receive instruction on what the target structure is, while implicit instruction refers to a situation in which learners do not receive such instruction (R. Ellis, 2009). Hypothetically, explicit instruction could lead to implicit learning and result in implicit knowledge, for example, when learners are instructed on what they have to learn in the second language they are learning, but cannot understand this instruction and end up with knowledge they are not aware of. Yet, from empirical studies we know that explicit instruction often leads to a higher degree of explicit learning, and results in more explicit knowledge (e.g. Hamrick & Rebuschat, 2012), and the same holds for implicit instruction, learning and knowledge (e.g. Williams, 2005). Still, it is important to keep in mind that neither explicit learning, knowledge and instruction nor implicit learning, knowledge and instruction necessarily need to co-occur (Batstone, 2002; DeKeyser, 2009, Williams, 2009).

5.2 Studies on explicit and implicit learning

Many studies support the idea that explicit learning mechanisms play a role in adult language learning and that implicit learning mechanisms contribute to child language learning. Adult learners are often greatly helped by explicit instruction, as evidenced by studies that show that explicit instruction has a positive effect on learning outcomes (for meta-analyses see Goo, Granena, Yilmaz & Novella, 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). Meta-analyses show that adults who receive explicit instruction, whether they are explicitly instructed on pragmatics (e.g. Alcón Soler, 2005; Fukuya & Martinez-Flor, 2008; Koike & Pearson, 2005; Martinez-Flor & Fukuya, 2005; Muranoi, 2000; Sheen, 2007), syntax (e.g. Gass, Svetics & Lemelin, 2003; Robinson, 1997a; 1997b; Scott, 1989), morpho-syntax (e.g. DeKeyser, 1995; Ellis, Loewen & Erlam, 2006; Gass, et al., 2003; Williams & Evans, 1998) or phonetics (e.g. Kissling, 2013; Saito, 2013), often show higher learning gains than uninstructed language learners in classroom or experimental settings.

Evidence from the statistical learning literature suggests that children make use of implicit learning mechanisms to acquire fundamental parts of language (e.g. Erickson & Thiessen, 2015; Frost & Monaghan, 2016; Romberg & Saffran, 2010). Statistical learning is defined as the ability to implicitly extract linguistic regularities from distributional properties in the input (e.g. Arciuli & Simpson, 2012; Aslin et al., 1998; Baker et al., 2004; Conway et al., 2010; Conway

& Christiansen, 2006; Kidd, 2011; Perruchet & Pacton, 2006; Reber, 1967; Rebuschat, 2013; Rebuschat & Williams, 2012; Turk-Browne et al., 2005). This learning mechanism enables both children and adults to infer word boundaries (e.g. for children Aslin et al., 1998; Gómez, 2002; Gomez & Gerken, 1999; Saffran, Johnson & Aslin, 1996; for adults Endress & Bonatti, 2007; Perruchet & Vinter, 1998; Thiessen et al., 2013), acquire words and their meanings (e.g. for children Smith et al., 2014; Smith & Yu, 2008; Vlach & Johnson, 2013; for adults Kachergis et al., 2014; Vouloumanos, 2008; Yu & Smith, 2007), and form grammatical categories (e.g. for children Lany, 2014; Lany & Saffran, 2013; for adults Chen et al., 2017; Monaghan et al., 2015). Taken together, and under the assumption that statistical learning really is implicit, these statistical learning studies and the earlier described instruction studies seem to show that children learn implicitly, and that adults, in addition to learning implicitly, can learn explicitly as well. However, we know very little about the role of explicit learning processes when children learn grammatical rules, because this has hardly ever been researched.

There is some work on the role of explicit instruction in young children. Several studies have investigated the role of explicit instruction in the acquisition of vocabulary items in this population (e.g. Coyne et al., 2009; Silverman, 2007; Vaahtoranta et al., 2018). A meta-analysis of such studies shows that vocabulary training in kindergarteners is effective and that more explicit instruction conditions result in bigger effect sizes (Marulis & Neuman, 2010). Furthermore, the way in which explicit instruction in phonemic awareness relates to reading and spelling abilities has been widely studied in similar age groups as well (e.g. Degé & Schwarzer, 2011; Furnes & Samuelsson, 2011; Gillon, 2018; Johnson & Goswami, 2010; Saygin et al. 2013). A meta-analysis by Ehri et al. (2011) on this topic finds that instruction aimed at phonemic awareness significantly aids both reading and spelling in preschoolers, kindergarteners, and first graders. The work on explicit instruction in children, although limited to the domains of vocabulary, and reading and spelling, indicates that young learners can benefit from receiving explicit instruction when acquiring particular linguistic knowledge. However, whether explicit instruction also influences learning grammatical structures in these age groups, is yet an unresolved question.

Lichtman (2016) addressed this question by investigating the effect of explicit instruction in a group of children from 5 to 7 years old, and compared this to adult learners. Participants learned a new language 'Sillyspeak' over six consecutive days, in which they had to acquire a particular word order and a determiner system. During the six days, half of the participants received only exposure to this language, and the other half received an explicit training session instead of a part of the exposure. Participants who received explicit instruction were taught grammatical gender on the determiners using pictures, and were taught word order by contrasting it to the English system. On the seventh day, they were tested on their knowledge of the language. The study did not show an effect of explicit instruction in children on accuracy when producing these grammatical structures, but also did not show such an effect in the adult group. Crucially, because explicitly instructed participants received less exposure than uninstructed peers, it is difficult to determine whether explicit instruction indeed does not make a difference for children. Perhaps explicit instruction helped these young learners to compensate for the extra exposure that uninstructed learners received. Alternatively, explicit instruction may not have added value to the input children received over the course of six days, even though this input was not equal for both groups. The imbalance of exposure seems to make it difficult to identify the actual effect of explicit instruction, which is not an uncommon problem in instruction studies (Andringa et al., 2011). These results thus warrant further investigation. Do young children benefit from explicit instruction, or does it make no difference whatsoever?

To answer to this question, we will delve deeper into this issue and put to test claims as put forward by Bialystok (1994), which others followed up upon (DeKeyser & Larson-Hall, 2005; Paradis, 2004, 2009), that children learn grammatical regularities only implicitly, and do not make use of explicit learning mechanisms. Using the same miniature language learning experiment as in Chapters 2 and 4, we investigated whether young learners who receive both exposure and explicit instruction demonstrate better learning of a morpho-syntactic element than young learners who receive the same amount of exposure but without explicit instruction. We were thus not interested in whether or not children use explicit instruction differently than adults, but merely whether they use it at all.

5.3 Current Study

To test if children can make use of explicit learning mechanisms through explicit instruction, we investigated whether explicit instruction influences the acquisition of a number marker in kindergarteners. We used the artificial language learning design from Chapters 2 and 4, in which we tested whether kindergarteners could learn a number marker from mere exposure. In the current experiment we compared kindergarteners who had to learn this marker only from exposure with kindergarteners who were exposed to the input in addition to receiving explicit instruction about the number marker. The studies in Chapters 2 and 4 probed into children's knowledge using only a picture matching task, a technique that is considered a more explicit measure in second language acquisition. This study also followed their eye movements during the task to gauge knowledge, which is commonly regarded as a more implicit measure. Apart from the eye tracking measure, the present experiment was identical to our previous experiments for children in the exposure only group. As such, the current study provides another opportunity to partially replicate our previous findings (see e.g. Marsden et al., 2018 for a discussion about replications in linguistics).

As we briefly mentioned in Chapter 1, it is important to include measures of both explicit and implicit knowledge when investigating the possible effects of explicit instruction on learning a grammatical element (Andringa et al., 2011; Andringa & Rebuschat, 2015; R. Ellis, 2005; Han & Ellis, 1998). Explicit instruction often leads to a higher degree of explicit learning and results in explicit knowledge (e.g. Hamrick & Rebuschat, 2012; Williams, 2005), but explicit instruction could also lead to implicit knowledge (Batstone, 2002; DeKeyser, 2009, Williams, 2009). To measure learning, researchers have used grammaticality judgement tasks (Bader & Häussler, 2010; Schütze & Sprouse, 2013), picture matching tasks (Pan, 2012; Schmidt & Miller, 2010), eye tracking paradigms (Sagarra & Seibert Hanson, 2011; Sedivy, 2010), reaction times (Kaiser, 2013; Marinis, 2010), and other experimental methodologies (Blom & Unsworth, 2010; Hoff, 2012; Podesva & Sharma, 2013). None of these methods necessarily reveal anything about the implicit or explicit status of the knowledge, but each is more strongly associated with one type of knowledge or the other (Andringa & Rebuschat, 2015; Bowles, 2011; R. Ellis, 2005; Godfroid et al., 2015; Han & Ellis, 1998).

On the one hand, measures like timed judgement tasks, eye tracking paradigms and reaction time studies have been claimed to measure implicit knowledge, whereas on the other hand, untimed judgement tasks, meta-linguistic knowledge tasks, and picture matching tasks are often more associated with explicit knowledge (Andringa & Rebuschat, 2015; Godfroid et al., 2015). The former types of measures are performed under time pressure or do not explicitly ask participants something about the grammaticality of test items. This makes it unlikely that participants consciously steer their behavior during the task, and that explicit knowledge drives task performance (R. Ellis, 2005; Loewen, 2009). The latter types of measures often focus on the form and grammaticality of linguistic stimuli very explicitly. This makes it more likely that explicit knowledge drives task behavior of the participant, although implicit knowledge could still be involved during these tasks as well (R. Ellis, 2005; Sagarra & Seibert Hanson, 2011). Importantly, to get a better picture of the exact influence explicit instruction might have on learning, we should include measures that are associated with both implicit and explicit knowledge, as explicit instruction potentially influences both types of knowledge (Godfroid, 2015).

5.4 Method

5.4.1 Participants

113 Dutch speaking children (59 males, 54 females, *M* = 5;6 years, *SD* = 0;9) took part in this experiment. All children were in kindergarten and were recruited from four primary schools in the western and eastern regions of the Netherlands. They did not have any diagnosed language or communication disorders. No restrictions were imposed upon taking part in the experiment, and no further background information was collected. 10 participants had to be excluded, because of problems with calibrating (*N* = 2), other technical issues (*N* = 7) or because they did not want to participate anymore (*N* = 1). As a result, we present data from 103 Dutch speaking children (52 males, 51 females, *M* = 5;7 years, *SD* = 0;8). From this group, 50 children received explicit instruction during the training (24 males, 26 females, *M* = 5;7 years, *SD* = 0;8), the other 53 children did not receive such instruction and were only exposed to the miniature language (28 males, 25 females, *M* = 5;6 years, *SD* = 0;9). In the remainder we refer to the former group as the explicitly instructed group and to the latter as learners who received only exposure.

5.4.2 Materials

We used the artificial language from Chapters 2 and 4, which consisted of four proper names, three verbs, two grammatical markers, six frequent nouns, twelve infrequent nouns and one conjunction. Apart from the proper names, all words were novel. The language had subject-verb-object word order. Recall from the discussion in previous chapters that, in this language, a noun phrase on its own does not encode number and could correspond to both singular and plural referents. In a sentence, however, an argument noun phrase must be introduced by a nominal marker (*pli* or *tra*). In our miniature language, a noun introduced by *tra* can refer to both singular and plural referents: the correct interpretation must be inferred from the visual context. The language includes another nominal grammatical marker *pli*, which encodes number and indicates that the noun necessarily refers to multiple referents. In short, the rule participants had to learn was that whenever *pli* preceded a noun, this noun always referred to multiple referents.

5.4.3 Structure of the experiment

The overall structure of the experiment was very similar to the experiment described in Chapters 2 and 4, and which is illustrated by Figure 5.1. Since this experiment is to a great extent similar to that in the previous chapters, we refer the reader to those chapters for more detailed information. To learn the regularity, participants were told a story with four protagonists (Carlo, Julia, Marco and Maria), who were going on a holiday to a country where they did not speak the language. Participants were asked to help the protagonists learn this new language. They were told they would see pictures and hear things in the new language that matched the pictures they saw. The experiment consisted of four parts. It started out with a small vocabulary training. After the vocabulary training, a rule training session followed. During this session, children received input to acquire the grammatical rule.

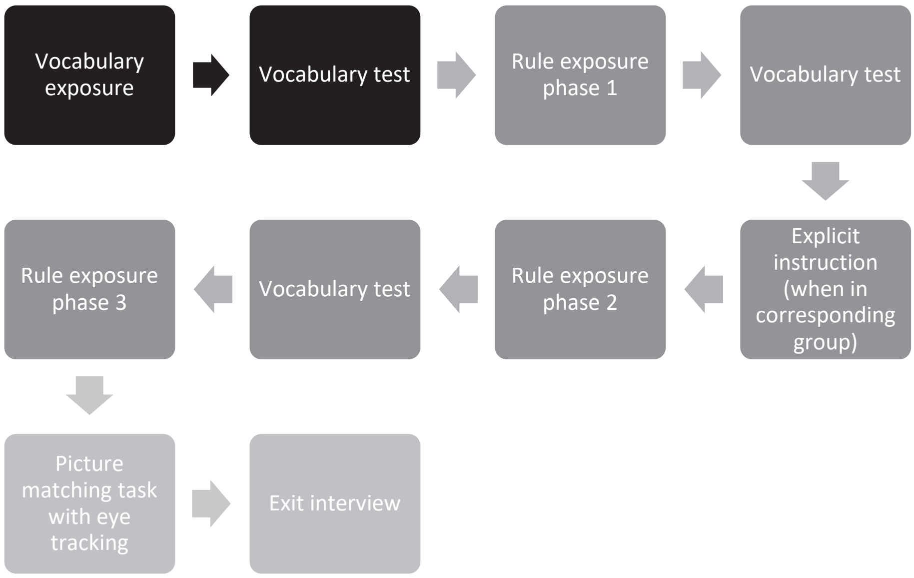


Figure 5.1. A visual representation of the structure of the experiment. The darkest boxes show the parts of the vocabulary training. The lighter boxes show the different parts of the rule training. The lightest boxes show the test phase.

Children who were in the explicit instruction group received instruction immediately before the second rule learning phase, after they were exposed to 1/3 of the training sentences. During this instruction they were told the meaning of both grammatical markers *pli* and *tra* by one of the characters in the game. The Dutch instruction and its English translation can be seen in (5) and (6). The instruction was presented in smaller segments, which are indicated by dashes. Children would also see pictures accompanying parts of the instruction. When they would hear that a noun introduced by *pli* referred to two or three objects, they would see pictures with two or three objects respectively. Segments combined with an accompanying picture are in bold face.

(5) "Je ziet straks steeds weer een plaatje, en je hoort iets in de nieuwe taal. // Als je goed luistert, hoor je in de taal soms het woordje *tra*. Heb je dat al eens gehoord? Luister maar. // *Maria rigarda tra herbo* // Elke keer als je *tra* hoort zie je op het plaatje // **één** // **twee** // **of drie dingen** // **Kijk maar.** // Misschien heb je soms ook wel het woordje *pli* gehoord. Luister maar. //*Carlo rigarda pli zambo*. // Als je *pli* hoort staan er altijd // **twee** // **of drie dingen op het plaatje**. // **Zie je dat?** // Zullen we nu weer verder gaan?

(6) "You will see some pictures again, and will listen to the new language. // If you listen carefully, you will sometimes hear the word *tra* in the language. Did you hear that? Listen. // *Maria rigarda tra herbo*. // Every time you hear *tra* you will see // **one** // **two** // **or three things on the picture** // **See**? // Sometimes you might have heard the word *pli*. Listen. // *Carlo rigarda pli zambo*. // Every time you hear *pli* there are always // **two** // **or three things on the picture.** // **See?** // Shall we continue?"

After the rule training, participants performed a picture matching task to test whether they acquired the rule. In the picture matching task, participants heard sentences from the rule training phase, and had to choose which of two pictures matched each sentence. For the test items of this task, participants had to choose between pictures with either one or multiple referents. The two pictures always referred to different referents. 12 of the experimental items contained *pli*. For these items, the target picture always showed multiple referents. The other 12 of the experimental items contained *tra*. Any number of referents would be grammatical for these items. Sensitivity to the statistical regularity in the input would lead to more target answers on trials with *pli* than on trials with *tra*, because for trials with *pli* participants could base their response on the number of referents the picture showed, whereas the number of referents was not indicative for trials with *tra*.

To test whether explicit instruction also influenced more implicit knowledge, we measured participants' eye movements during the picture matching task. Participants could see the two pictures from the onset of the sentence, but we were interested at where they would look from the onset of the grammatical markers *pli* and *tra*, as this was the point at which we could expect eye movements to differ for the two conditions. Participants could look more often towards the target picture after the onset of *pli*, as this marker predicted the correct answer, whereas participants should not have a looking preference for any of the two pictures after the onset of *tra*, as this did not predict the correct answer. If participants acquired the regularity, this would lead towards more correct looks towards the target after the onset of the grammatical marker for sentences with *pli,* than for sentences with *tra*. If explicit instruction influences online processing in some way, a possible looking preference for the target picture for sentences with *pli* over sentences with *tra* might be observed at a different point in time in the explicitly instructed group than in the group that received exposure only.

We presented the two pictures between which participants had to choose during the picture matching task in colored frames. Participants had to name the color of the frame of the correct picture. We used eight basic colors (black, blue, green, orange, pink, purple, red, yellow) to create these frames, which were semi-randomly distributed over all items. Color combinations were chosen such that the two colors contrasted with each other, to help participants distinguish between the colors they had to name. We used the color frames to minimize possible data loss during the task, as there are many threats to gathering eye tracking data in young participants (Léger et al., 2018). This way, participants did not have to point towards the correct picture, which might interfere with their eye movement measurements. They also did not have to press a button corresponding to the correct picture, which could have caused children to look at the buttons they need to press instead of the screen. Because of possible difficulties with the distinction between left and right in this age group, we could also not let them call out the correct picture on the basis of the side of the screen.

After the picture matching task, an informal debriefing took place in which we examined whether participants were aware of their knowledge in a way they could verbalize. During this debriefing, participants were asked how they knew what the correct answer was during the picture matching task, and whether they knew the meaning of the words *pli* and *tra*. To summarize, this means that apart from the following points, the present and previous experiments were exactly the same, and all statistical procedures to analyze the data from the picture matching task in the current experiment were identical to the procedures we followed in our previous experiment.

1. The vocabulary test after the vocabulary training at the start of the present experiment contained only two test items per noun. In Chapter 2, all participants continued to be exposed to test items until they identified all six nouns correctly four times. This proved frustrating, however, because most children had difficulties learning all nouns from the limited input they received. Therefore, it was problematic to keep the vocabulary training as it was, and it was altered from the test in Chapter 4, as it is in the present study. The vocabulary test was included to familiarize participants with the picture matching task, and participants did not receive any input to the grammatical regularity at this point. We think it is unlikely that this change would affect learning of the grammatical regularity.

2. All participants in the present study performed the attention task, whereas only half of the participants performed this task in Chapter 2. In the experiment presented in that chapter, we observed a tendency towards better performance by children in the attention task condition. Therefore, we decided to include the attention task for all children in Chapter 4 and the current study.
3. In Chapter 4, the experiment ended with an opt out phase, in which we assessed whether children developed awareness of the grammatical regularity. It followed the picture matching task, and could not have affected performance on this task. The opt out phase took place before the informal debriefing, and could have influenced the answers children gave during this debriefing in that study. This opt out phase was not part of the experiment from Chapter 2, and the current experiment.
4. In the present study, we measured eye movements during the final picture matching task, which were not measured in the studies from Chapters 2 and 4. This means that the current picture matching task was slightly different from the previous two studies. In the current task, there was a small pause between items in which participants had to fixate on a cross in the middle of the screen. Also, in the previous studies, pictures were not presented in colored frames, which were necessary to enable children to give an answer without pointing to the screen. In Chapters 2 and 4, children could simply point towards the screen, which might have been cognitively less demanding.

5.4.4 Procedure

As in previous experiments reported here, all stimuli were recorded by a female native speaker of Dutch. Sentences were recorded as a whole, leading to minor natural variation between similar words in the language. The test was administered in a quiet room at the participants' school. The task was presented on a laptop using E-prime (Psychology Software Tools, 2012). Eye tracking data was gathered using a Tobii Pro X3-120 Eye Tracker with an External Processing Unit. During vocabulary training, nouns and their accompanying pictures were presented for three seconds, before automatically moving to the next noun. As the goals of this training were familiarization and keeping attention, we did not register scores during the vocabulary test systematically enough to report on. During rule training, sentences and their accompanying pictures were presented for four seconds, before automatically moving to the next sentence. The instruction was prerecorded and segmented into smaller bits, such that the instructor could press a button on the keyboard to continue with the next piece of instruction. During the test phase, the experimenter pressed a button on the keyboard that corresponded to the answer the participant gave. Scores from this phase were registered automatically. We calibrated the eye tracker before the start of the test phase, and recorded eye movements at a rate of 120 Hz. After each test item, participants saw a fixation cross in the middle of the screen. The following test item started automatically, if participants fixated for 100 ms (12 samples) on the cross. This ensured that participants were looking at the middle of the screen, between the pictures, at the onset of every test item. For every test item, the audio and pictures were presented simultaneously. Both the participant and the experimenter listened to the audio using headphones. The vocabulary training lasted 8 minutes on average, the rule training phase 15 minutes, and the picture matching task including calibration 10 minutes. The full experiment took approximately 30 minutes per participant. Ethical approval for this study was obtained from the University of Amsterdam and passive consent was obtained from children's parents or legal guardians before the start of the study.

5.4.5 Analysis

All analyses were carried out in R (R Core Team, 2015) using the lme4 package (Bates et al., 2015) and eyetrackingR package (Dink & Ferguson, 2015) where needed. To determine whether participants grasped the target regularity, and whether explicit instruction affected this learning, two separate analyses were carried out. We used linear mixed effects modeling to analyze the results of the picture matching task, and a cluster based permutation analysis to analyze the eye tracking data. We also carried out an additional cluster based permutation analysis on filler items, to test whether our groups were comparable in their referential processing.

To investigate the effect of explicit instruction on learning as measured by the picture matching task, a generalized linear regression model with mixed effects was carried out to investigate the results from the picture matching task. This model took the responses from the task (1 for a correct answer, 0 for an incorrect answer) as a dependent variable, marker type as a within-participants fixed effect, type of instruction as a between-participants fixed effect, participant as a between-participants random effect and item as a within-participants random effect. We did not include any random slopes in our model. Our fixed effects were included in this model because we were *a priori* interested in their contribution to the outcome (Gelman & Hill, 2007). Additionally, as this model does not provide an outcome for the learning effect in just the exposure only group, we ran another generalized linear regression model in this group only to see whether we replicated the findings from our previous studies. This model took the responses from the task (1 for a correct answer, 0 for an incorrect answer) as a dependent variable, marker type as a within-participants fixed effect, participant as a between-participants random effect and item as a within-participants random effect.

In both models, orthogonal sum-to-zero contrast coding was applied to our binary fixed effects (i.e., marker type and instruction type; Baguley, 2012, p590-621). We aimed to keep both models as fully specified as possible by including random intercepts for participants and items (Barr et al., 2013). We increased the number of possible iterations to 100.000 (Powell, 2009), to solve potential issues with non-converging models. This enabled us to report on a random effect structure justified by our data (Jaeger, 2009). We report simple rather than standardized effect sizes (Baguley, 2009) and Wald confidence intervals (Agresti & Coull, 1998).

To investigate whether explicit instruction influenced looking behavior during the picture matching task, we carried out a cluster based permutation analysis (see Curcic, Andringa & Kuiken, 2019; Dink & Ferguson, 2015; Maris & Oostenveld 2007) using the eyetrackingR package (Dink & Ferguson 2015). This analysis enabled us to investigate the development of the proportion of looks towards the target picture over time. We analyzed trials from 500 ms before the onset of the grammatical marker until 4000 ms after the onset of the marker (see Figures 5.3-5.6 for a visual illustration of this time window). On average a grammatical marker lasted 350 ms, there was approximately 200 ms between the offset of the marker and the onset of the noun, and a noun lasted about 550 ms. A sentence thus took around 1100 ms from the onset of the marker. This means the endpoint of the frame we analyzed was approximately 3000 ms after the offset of the sentence. As children often took at least 3000 ms to give an answer during the task, we have eye tracking data for the majority of the test items in this timeframe. After this point in time, we have fewer data points available, as children might have given an answer, and the next test item would be played. When interpreting the eye tracking data, it is important to know that in this age group it takes approximately 400 ms before eye movements reflect the processing of a presented stimulus (Clackson, Felser & Clashen, 2011; Fukushima, Hatta & Fukushima, 2000).

For the cluster based permutation analysis, we analyzed all test items, regardless of whether participants gave a correct or incorrect answer during the picture matching task. For every item, we coded whether a single look was directed at the target picture (1) or not (0). The area of interest for both types of pictures was the whole area within the colored frames. We removed all looks that were irrelevant (14.11%), i.e., on screen but not on either of the two pictures, or missing (12.55%). After removing looks that were classified as either irrelevant or missing, we excluded trials in which more than 50% of the looks had been removed as a result of this procedure (9.1% of the total number of test items). This analysis, which involved two steps, allowed us to determine at which parts within the marked time frame, participants were likely to look more often towards the target picture for sentences with *pli* than for sentences with *tra*. In the first step, the data were split into 50 ms (6 samples) time bins, and by calculating a *t*-value for each time bin we determined whether participants looked significantly more often towards the target picture for *pli* sentences than for *tra* sentences. Adjacent time bins for which there was a significant effect were clustered together, and the *t*-values were summed. In a second step, the data were reshuffled hundreds of times, to see how likely these clusters and accompanying summed statistics would be if the data were randomly generated. This way, *p*-values could be calculated which indicate how likely the observed clusters and summed *t*-values would be observed if the data were generated randomly.

Using this procedure, we can correct for false alarms, while not losing much of the sensitivity of the data. The expectation was that children who acquired the number marker would be more likely to look at the correct image in *pli* sentences than in *tra* sentences. Because we expected the effect to occur in this direction, we used a cut-off point of 0.05 for significance. Unfortunately, a cluster based permutation analysis only allows for one predictor and was thus unsuitable for an analysis in which we looked at the effect of marker and instruction type simultaneously.[2] Therefore, the data for our explicitly instructed group and the uninstructed group were analyzed separately. We had no specific expectations about the effect of instruction on eye movement behavior. Possible outcomes could be that instruction type would lead to predictive eye movements in one group, but not in the other, or that such prediction would occur at different moments in time.

As a check on whether referential processing was similar in both groups, we used a similar procedure to analyze filler items. For this cluster based permutation analysis, we analyzed all fillers, regardless of whether participants gave a correct or incorrect answer during the picture matching task. For every item, we coded whether a single look was directed at the target picture (1) or not (0). We removed all looks that were irrelevant (14.21%), i.e., on screen but not on either of the two pictures, or missing (11.23%). After removing looks that were classified as either irrelevant or missing, we excluded trials in which more than 50% of the looks had been removed as a result of this procedure (4.61% of the total number of fillers). However instead of comparing looks between *pli* and *tra* sentences, we here compared the proportion of looks towards the target for fillers sentences between instructed children and uninstructed children. If our groups were comparable in their general referential processing, we should not observe any differences between groups with regard to their eye movements during these filler sentences.

---

[2] Chan, Yang, Chang & Kidd (2018) for example compare a cluster based permutation analysis, similar to our analysis, to a regression model in which more than one predictor can be taken into account, but they argue the latter analysis is a weaker one for this type of data. Perhaps this is the case, because scores need to be aggregated to make the data suitable for a regression analysis.

5.5 Results

An overview of the descriptive statistics from the picture matching task can be found in Table 5.1 and Figure 5.2. For this task we investigated whether children gave more correct answers for sentences with *pli* than sentences with *tra*. This would reflect learning of the number marker. We also considered whether this behavior interacted with the type of instruction children received. The table shows that participants gave more target answers when sentences contained *pli* than when sentences contained *tra*, but this effect was not significant (OR = 1.236, 95% CI = [0.873, 1.750], z = 1.194, p = .233). Results also show that explicitly instructed learners gave slightly less target answers in general than learners who received only exposure, but this effect was not significant either (OR = 0.938, 95% CI = [0.774, 1.136], z = -0.658, p = .511). Additionally, we observed a non-significant interaction between type of instruction and number marker (OR = 0.855, 95% CI = [0.618, 1.183], z = -0.944, p = .345). Although explicitly instructed learners seemingly showed less learning than their peers who received only exposure, this effect was only small and not significant. Furthermore, the mixed-effects model that was conducted to investigate learning in just the exposure only group, showed learners who received just exposure gave more correct answers for sentences with *pli* than for sentences containing *tra* (OR = 1.334, 95% CI = [0.913, 1.948], z = 1.490, p = .136). Again, this effect was not significant.

Table 5.1. Scores from the picture matching task.

	Explicit Instruction ($N = 50$)			Exposure only ($N = 53$)		
	M	**SD**	**Range**	**M**	**SD**	**Range**
Pli	6.40	2.07	1-11	6.79	1.78	3-11
Tra	6.02	1.91	0-10	5.98	1.79	2-10

Note: Scores indicate the number of target answers produced by children who received explicit instruction and exposure, and children who received only exposure. Sentences with grammatical marker *pli* were predictable, which should lead to more correct target answers in the case of learning. Scores could range from 0-12.

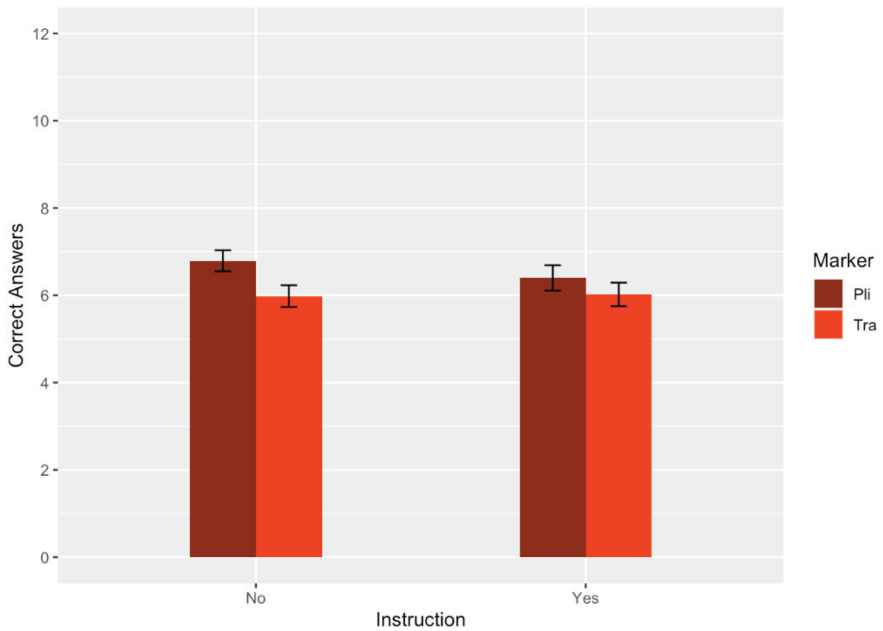


Figure 5.2 A graph depicting the results from the picture matching task for uninstructed (left) and explicitly instructed (right) learners. Scores from the picture matching task indicated the number of target answers. Scores could range from 0-12.

As a first step in the analysis of the eye tracking data, we looked at the explicitly instructed learners and learners who received only exposure together, to see whether their eye movements revealed sensitivity to the grammatical rule. To test this, we conducted a cluster based permutation analysis with the proportion of looks towards the target picture as a dependent variable, and marker type as a within-participants factor. Figure 5.3 and Table 5.2 show the results of this analysis. In the first step of the analysis, we observed that participants behaved differently for sentences with *pli* than for sentences with *tra* during five clusters of time bins. In these cases, positive sum statistics were observed indicating that participants were more likely to look towards the target picture in *pli* sentences than in *tra* sentences. In the second step, we found that participants looked significantly more often towards the target picture for sentences with *pli* than for sentences with *tra* in the cluster that lasted 1150 ms, and started at 700 ms after the onset of the marker till 1850 after this onset. This cluster starts 150 ms after the onset of the noun and lasts till approximately 750 ms after the offset of the noun, and thus the end of the sentence.

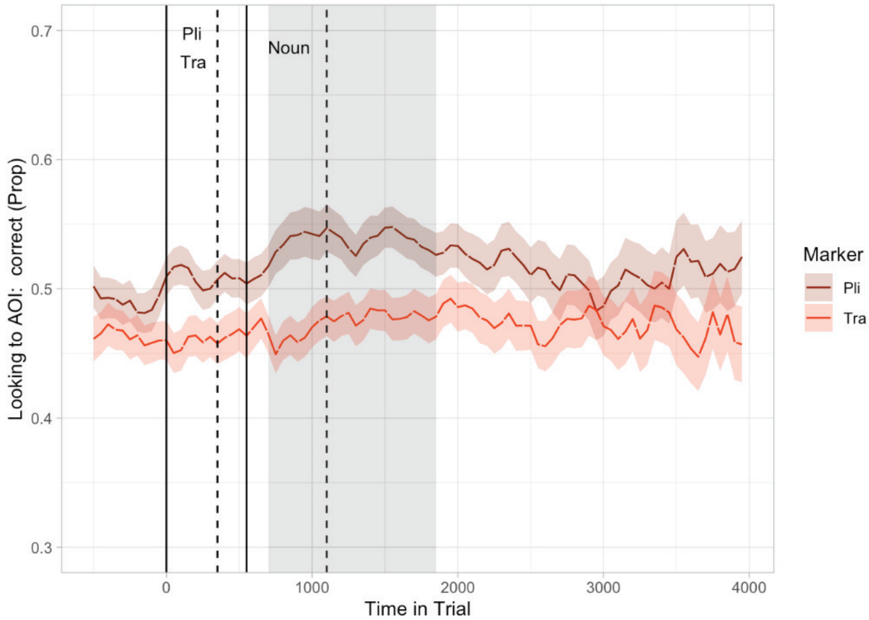


Figure 5.3. Proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra*. Proportion of looks was calculated from all looks that were either on the target or on the non-target picture. Solid lines indicate the onset of the grammatical marker (*pli* or *tra*) and the noun respectively, dashed lines the offsets of these words. The shaded area indicates the cluster of time bins during which an effect of grammatical marker was observed.

Table 5.2. Results of the cluster based permutation analysis for explicitly instructed learners and learners who received only exposure together.

Time cluster	Sum statistic	Time range (ms)	Probability
1	11.06	0 – 200	.264
2	6.74	350 – 500	.417
3	**64.75**	**700 – 1850**	**.001**
4	2.09	2550 – 2600	.672
5	2.04	3950 – 4000	.704

Note: The time range is measured from the onset of the grammatical marker. The cluster in bold turned out to be significant after the second step of the analysis.

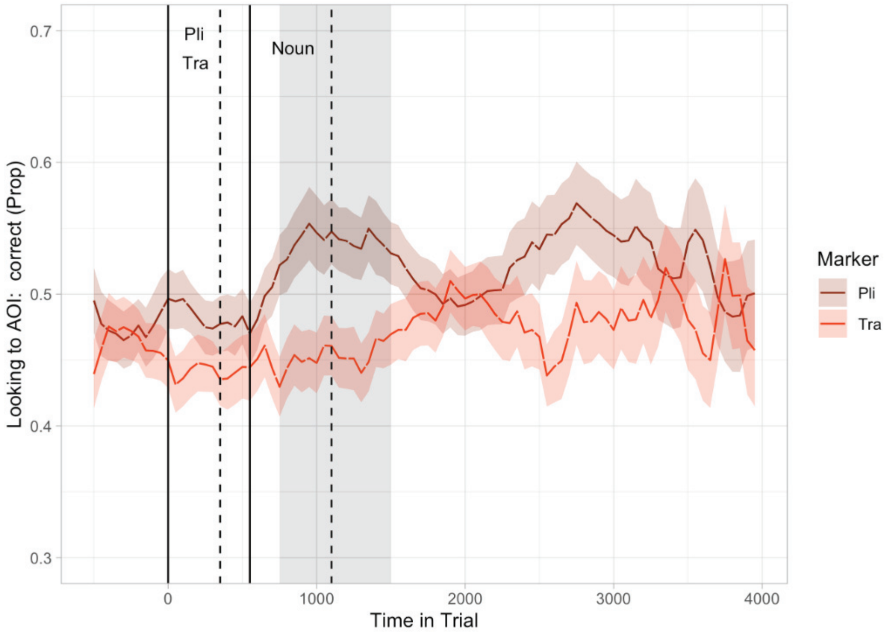


Figure 5.4. Proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra* for the explicitly instructed group. Proportion of looks was calculated from looks that were on the target or on the non-target picture. Solid lines indicate the onset of the grammatical marker (*pli* or *tra*) and the noun respectively, dashed lines the offsets of these words. The shaded area indicates the cluster of time bins during which an effect of grammatical marker was observed.

Table 5.3. Results of the cluster based permutation analysis for the explicitly instructed learners.

Time cluster	Sum statistic	Time range (ms)	Probability
1	4.48	50 – 150	.534
2	**41.62**	**750 – 1500**	**.009**
3	10.62	2550 – 2750	.266
4	2.22	2800 - 2850	.675

Note: The time range is measured from the onset of the grammatical marker. The cluster in bold turned out to be significant after the second step of the analysis.

This analysis was repeated for the explicitly instructed and exposure only groups separately. Figure 5.4 and Table 5.3 show the results of the analysis for the explicitly instructed group. In the first step of this analysis, we observed four clusters of time bins during which participants looked more often towards the target picture for sentences with *pli* than for sentences with *tra*. In the second step we found that this effect was only significant in the cluster that lasted from 750 ms after the onset of the marker till 1500 after this onset. This cluster starts 150 ms after the onset of the noun and lasts till approximately 400 ms after the offset of the noun, and thus the end of the sentence. Figure 5.5 and Table 5.4 show the results of the analysis for the exposure only group. In the first step of this analysis, we observed four clusters of time bins during which participants looked more often towards the target picture for sentences with *pli* than for sentences with *tra*. In the second step, we found this effect reached significance for the cluster that lasted from 1500 ms after the onset of the marker till 2050 after this onset. This cluster starts 400 ms after the offset of the noun and lasts till approximately 950 ms after the offset of the noun, and thus the end of the sentence. In both groups, learners showed looking behavior that might be associated with learning at some point in time. Crucially, the moment at which this behavior occurred seems earlier in time for the explicitly instructed group than for the exposure only group.

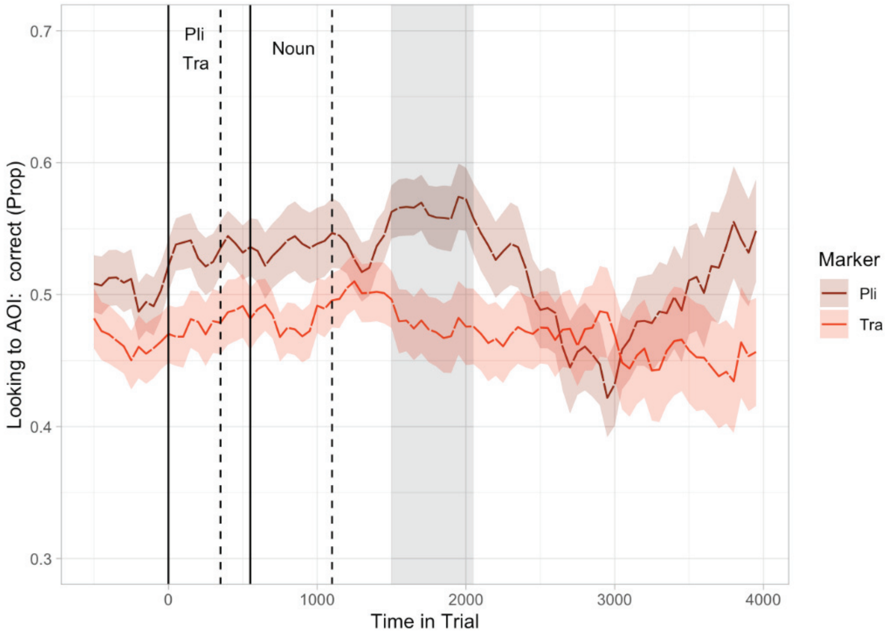


Figure 5.5. Proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra* for the exposure only group. Proportion of looks was calculated from looks that were on the target or on the non-target picture. Solid lines indicate the onset of the grammatical marker (*pli* or *tra*) and the noun respectively, dashed lines the offsets of these words. The shaded area indicates the cluster of time bins during which an effect of grammatical marker was observed.

Table 5.4. Results of the cluster based permutation analysis for the exposure only group.

Time cluster	Sum statistic	Time range (ms)	Probability
1	2.18	-250 – -200	.605
2	4.33	50 – 150	.496
3	**27.83**	**1500 – 2050**	**.053**
4	2.41	3800 – 3850	.555

Note: The time range is measured from the onset of the grammatical marker. The cluster in bold turned out to be nearly significant after the second step of the analysis.

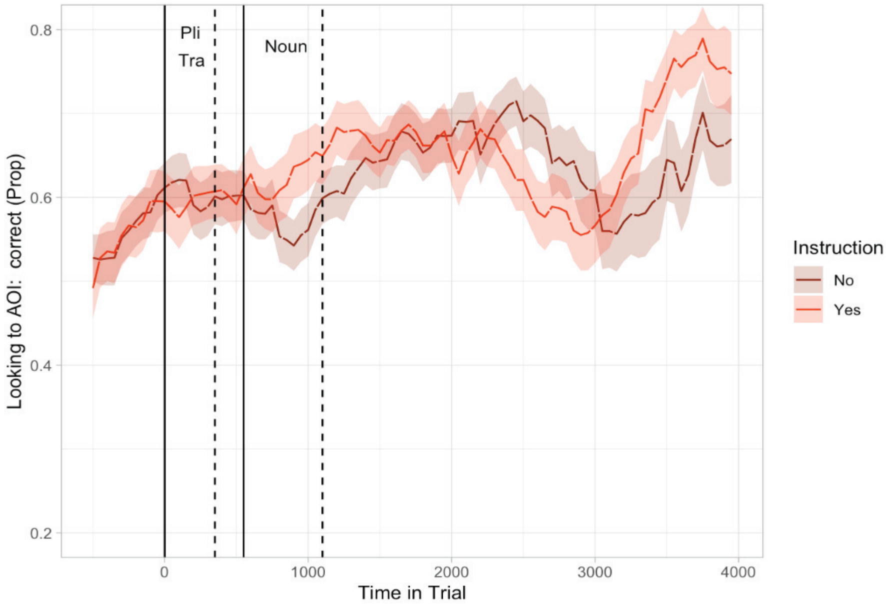


Figure 5.6. Proportion of looks towards the target picture over time for filler sentences for the explicitly instructed group and the exposure only group. Proportion of looks was calculated from all looks that were either on the target or on the non-target picture. Solid lines indicate the onset of the grammatical marker (*pli* or *tra*) and the noun respectively, dashed lines the offsets of these words.

Table 5.5. Results of the cluster based permutation analysis for the exposure only group.

Time cluster	Sum statistic	Time range (ms)	Probability
1	2.00	2450 - 2500	.726
2	4.10	2600 - 2700	.562
3	-6.75	3550 – 3700	.362

Note: The time range is measured from the onset of the grammatical marker.

To investigate whether the behavior we observed in our two groups can be attributed to a difference in received instruction, we checked whether referential processing differed between groups. Figure 5.6 and Table 5.5 show the results of the analysis. We carried out another cluster based analysis in which we compared looks towards the target picture in filler items between the instructed and uninstructed group. In the first step of this analysis, we observed three small clusters of time bins during which participants from one group looked more often towards the target picture for fillers than the other. In the second step, none of these three clusters turned out to be significant. Because these clusters were only small and non-significant, we assume that the differences in processing of the grammatical markers between groups are probably not caused by a general referential processing difference between instructed and uninstructed children.

In the debriefing, when children were asked how they made a decision on the picture matching task, participants either reported they did not know how they made a decision ($N = 83$) or they claimed that they had heard the sentences before and remembered what they meant ($N = 20$). When asked for the meaning of *pli* and *tra*, most children either said they did not know or could not give a meaning ($N = 81$), gave the meaning of a noun they learned during the exposure ($N = 13$), or some other meaning ($N = 7$). However, when asked what *pli* meant, one child in the exposure only group answered 'to calculate'; when asked what *tra* meant, one child in the explicitly instructed group answered 'one, two or three', which is a reasonable hypothesis in this case.

5.6 Discussion

At the beginning of the present chapter, we discussed how many scholars have claimed there is a fundamental difference between children and adults learning the grammatical structures of a new language (Bley-Vroman, 1990; Hartshorne et al., 2018; Krashen et al., 1982; Singleton & Ryan, 2004). Adults would rely mostly on explicit learning mechanisms, but can combine these with implicit learning mechanisms (DeKeyser, 2000, 2003; R. Ellis, 2005, 2009; Paradis, 2004, 2009; Ullman, 2001), whereas children would only use implicit learning mechanisms. This implies that children do not learn grammar explicitly, and that explicit instruction about the grammatical regularities in the target language does not enhance learning (Bialystok, 1994; DeKeyser & Larson-Hall, 2005; Paradis, 2004, 2009). The goal of this chapter was to investigate whether children indeed are not able to learn grammatical structures explicitly, by testing the effect of explicit instruction on the acquisition of a morphosyntactic element in kindergarteners. As explicit instruction could influence both more explicit and more implicit knowledge, we measured learning using a picture matching task and eye tracking, because these measures are associated more with explicit and implicit knowledge respectively (Andringa & Rebuschat, 2015; Godfroid et al., 2015). While investigating the influence of explicit instruction on acquiring a morphosyntactic element in kindergarteners, this study also enabled us to partially replicate our findings from studies presented in Chapter 2 and 4, in which we demonstrated that a number marker could be learned when children did not receive explicit instruction about this marker.

More evidence of learning the number marker would be provided if children in the exposure only group gave more correct answers during the picture matching task for sentences with the marker *pli*, as it was predictive of the target answer, than for sentences with the marker *tra*, which was not predictive of the target answer. When we isolated this group to allow for comparison with the studies from our previous chapters, we saw they showed this behavior, but the effect was not significant. Although these results could be seen as null results, which suggest that we were unable to replicate our previous findings, they are compatible with the outcomes of the two previous studies, which are listed in

Table 5.6. The effect size of the current study is similar to that of the previous two studies, and the effect sizes of the three studies fall within each other's confidence intervals. Taken together with the previous studies, we might think that these and our previous results show that kindergarteners can acquire this meaningful grammatical marker from mere exposure to (limited input). Perhaps it was more difficult for children to show their knowledge of the grammatical marker in this experiment because they had to answer more indirectly during the task. In our previous experiments children could point towards the screen to indicate their answer, whereas now they had to name the color of the frame, because pointing would interfere with their eye movements. Naming colors might be cognitively more demanding, and likely requires more explicit processing. As a result, the current task may tap more into the explicit knowledge that participants possibly develop during learning (Andringa & Rebuschat, 2015; Bowles, 2011; R. Ellis, 2005; Godfroid et al., 2015; Han & Ellis, 1998). Alternatively, it could be difficult to establish the learning effect in the picture matching with sufficient certainty in each individual study, because it is not particularly big.

Table 5.6. Results from the generalized logistic effects model for the exposure only group only of present experiment, and the two earlier studies that we were aiming to replicate with this part of the experiment.

Study	*OR*	95% CI	*z*	*p*
Chapter 2	1.515	1.071 - 2.257	2.178	.029
Chapter 4	1.344	0.980 - 1.843	1.834	.067
Current	1.334	0.913 - 1.948	1.490	.136

Replicating the learning effect in the exposure only group was not the main goal of this paper. We were primarily interested to see whether there would be a difference between the explicitly instructed group and the group that received exposure only. The interaction effect between learning and instruction type was only small and not significant, but the direction of the effect suggests that learning might be hindered by explicit instruction, as children who received rule explanation showed slightly less learning. Furthermore, the confidence interval of this interaction effect (95% CI = [0.612, 1.192]) suggests that an effect in the opposite direction could occur, but would likely not be very big either. Although we cannot interpret this null finding as indicating there is no effect of instruction type of learning, we might conclude it indicates that explicit instruction does not make a meaningful difference. Replication of this study, such as presented in the next chapter, is necessary to make stronger claims about the possible effect of explicit instruction on learning. Nevertheless, if the results of the present study are upheld, then that might mean that children learn more effectively implicitly, and do not make use of explicit learning mechanisms (Bialystok, 1994; DeKeyser & Larson-Hall, 2005; Paradis, 2004, 2009).

The eye tracking data seem to present a slightly different picture. When measuring the eye movements of the children during the picture matching task, we expected that if children learned the meaningful grammatical element this would lead to more looks towards the target for sentences with *pli* than for sentences with *tra* at some point in the sentence after the onset of these markers. Importantly, we had no specific expectations for the effect of instruction on this looking behavior. We observed the expected looking behavior in the group as a whole, as children looked more often at the target from 700 to 1850 ms after the onset of the marker. In interpreting these data, it is important to take into account that in kindergarteners it takes approximately 400 ms before eye movements reflect the processing of a presented stimulus (Clackson et al., 2011; Fukushima et al., 2000). Building on this, the present findings indicate that the start of the observed time cluster coincides with the offset of the grammatical marker, which suggests that it indeed triggered looks towards the target picture. The eye movements of all learners taken together seemingly reveal they develop some knowledge of the grammatical marker.

Interestingly, we observed this looking behavior at different moments in time for the explicitly instructed group and the exposure only group. Explicitly instructed learners looked significantly more often towards the target picture in sentences with *pli* from 750 to 1500 ms after the onset of the grammatical marker. When applying the 400 ms correction to these data, explicitly instructed learners' preference for looking at the correct target coincided with the offset of the grammatical marker, which is about the moment one can expect such looking behavior, till about the end of the sentence. This looking behavior emerged later in the learners who received exposure only, namely between 1500 and 2050 ms after the onset of the grammatical marker, and this effect approached significance (p = .053). Applying the 400 ms correction, we might say they started looking towards the correct picture in predictable sentences when the sentence has ended. Both groups seemed to show eye movements that are indicative of learning, but explicitly instructed children showed this earlier than their peers who merely received exposure. Although we should remain cautious in interpreting these results, especially because they are borderline significant for the exposure only group, it seems that the picture matching task accuracy data and eye tracking data stand in contrast to each other. While the accuracy data provide no evidence of an effect of explicit instruction on learning, the eye tracking data are incompatible with the idea the children do not make use of explicit learning mechanisms (Bialystok, 1994; DeKeyser & Larson-Hall, 2005; Paradis, 2004, 2009). These results might show that explicit instruction leads to slightly better developed knowledge of the acquired grammatical element, which results in faster processing of it. The question remains then, why we were unable to find a similar learning effect in the picture matching task.

One potential explanation of our findings is that exposure to the miniature language over this short period of time (only 30 minutes) leads to low level or unstable representations of the acquired linguistic knowledge. In addition, the instruction may have been too brief, and merely raised attention to the form of the markers, rather than to their function. Following this line of reasoning, we may have been unable to show an effect of explicit instruction on learning in the picture matching task because the children's linguistic knowledge was not developed strongly enough. It is possible that children need more exposure to develop stronger representations, or that if there is an effect of explicit instruction on acquisition, this becomes visible in accuracy over a longer period of time. There are good reasons to assume learning not only happens during training, but also afterwards when a learner is no longer involved in the learning task (Diekelmann & Born, 2010; Maquet, 2001; Nieuwenhuis, Folia, Forkstam, Jensen & Petersson, 2013; Stickgold, 2005). Knowledge that is initially stored in short term memory, needs to be strengthened and consolidated in long term memory (Dumay & Gaskell, 2007; Frost & Monaghan, 2017; Gómez, Bootzin & Nadel, 2006; Mirković & Gaskell, 2016; Williams & Horst, 2014). If we tested these children at a later moment in time or gave them more exposure to the language, we might be able to observe stronger patterns that become visible in the picture matching task as well. This could imply that the eye movements of the children in the current experiment are a precursor of knowledge of the grammatical marker that is still under development.

Alternatively, our results could point towards a difference in learning between children and adults. Although children can make use of explicit instruction, this does not result in more explicit knowledge, as we typically see in adults (e.g. Hamrick & Rebuschat, 2012; Morgan-Short, Steinhauer, Sanz & Ullmann, 2012; Williams, 2005). Some authors have claimed explicit instruction could also lead to implicit knowledge (Batstone, 2002; DeKeyser, 2009, Williams, 2009), and perhaps this applies to children in particular. The fundamental difference between children and adults then does not lie in the learning strategies available to these learners, but in the type of knowledge they develop. This might explain why we see a difference in their eye movements, which is associated more with implicit knowledge, but hardly any in the picture matching task, which is associated more with explicit knowledge. This latter finding could have been reinforced by the more indirect nature of the task in comparison to our two previous studies. As children had to respond by naming the colors of the frames, explicit processing might have played a bigger role, and possible effects of explicit instruction on this task were more difficult to observe because we were less likely to tap into implicit knowledge. Furthermore, it explains why only a single child was able to report on the meaning of one of the two markers during the debriefing. In this view, our findings entail that children are well able to learn explicitly, but unlikely to develop explicit knowledge. Rather, children only develop implicit knowledge, and maybe even more when explicitly instructed. This might occur when children know they are learning the marker, because the instruction raised their attention of it, but they nevertheless only develop implicit knowledge. If this is indeed the case, such a developmental pattern can only be observed when knowledge is measured appropriately. Possibly, children can use explicit learning mechanisms, but because they learn language in completely different contexts than adults, do not use them very often (Lichtman, 2016). However, to justify such claims, a direct comparison between children and adults would be necessary, as we do not know whether adults would develop more explicit knowledge during this experiment.

The way we just characterized the fundamental difference between child and adult language learning relies on a dichotomous distinction between implicit and explicit knowledge. It assumes that when children cannot verbalize their knowledge, they are not aware of it and the knowledge is therefore implicit. However, awareness likely exists in gradations (Allport, 1988; Cleeremans, 2008, 2011, 2014; Dehaene et al., 2017; Dennett, 1993; James, 1890), and learners can also be aware of knowledge despite the fact that they cannot reflect on it verbally.

Perhaps children develop knowledge that is explicit and available to awareness, but at a lower, non-verbalizable level. Possibly, we did not tap into implicit knowledge with the eye tracking measure, because we measured eye movements *during* the picture matching task, which they had to carry out quite consciously. Following this rationale, we might argue that the earlier predictive eye movements of the instructed children indicate increased awareness of the target regularity and development of some explicit knowledge. Yet, this knowledge is not available to children's awareness to be beneficial when they have to make a more conscious decision. As a result, explicit instruction did not lead to a meaningful difference in accuracy during the picture matching task. This would mean that a possible effect of explicit instruction is not fundamentally different in children and adults, but only manifests itself gradually. In both groups explicit instruction may lead to an increase in explicit knowledge, but children lack the ability to verbalize this acquired knowledge. Future research could include more fine-grained measures of awareness, such as the opt-out procedure from Chapters 3 and 4, to investigate whether explicit instruction in young children indeed leads to an increase in knowledge that is explicit in nature, but on which they simply cannot reflect verbally.

Although many authors suggest children do not learn language explicitly, such claims have rarely been tested experimentally. In this chapter, we used a miniature language learning experiment to test whether children can learn explicitly, by investigating if explicit instruction influences kindergarteners' learning of a grammatical marker. In a picture matching task, we found no clear evidence of an effect of explicit instruction on accuracy, but there was an effect on children's eye movements. This shows that explicit instruction influences learning in a young age group to some extent, but the interpretation of these effects remains an open question. Perhaps the effect of explicit instruction in this age group becomes clear only after a longer period of exposure time or a consolidation period, or children develop a different type of knowledge than adults when explicitly instructed. In the next chapter, we present an experiment that exactly aimed to do this, in order to get more insight into the exact effect that explicit instruction has on learning in this age group.

Chapter 6

The effects of instruction, input and sleep*

In Chapters 1 and 5 we discussed how many scholars assume implicit learning mechanisms play an important role in language acquisition (Bialystok, 1994; DeKeyser & Larson-Hall, 2005; Paradis, 2004, 2009; Ullman, 2001, 2016). Learning is implicit when learners do not know they are learning and do not have such intentions (e.g. Dörnyei, 2009; N. Ellis, 2015; Hulstijn, 2015). One type of learning that possibly relies on implicit learning mechanisms is statistical learning (Christiansen, 2019). This label refers to the mechanism with which learners detect regularities within the environment, for instance language (e.g., Erickson & Thiessen, 2015; Frost & Monaghan, 2016; Romberg & Saffran, 2010, and see also Chapters 1 and 2 for further discussion). Statistical learning is often assumed to be an implicit process, happening without any awareness involved (e.g., Arciuli & Simpson, 2012; Aslin et al., 1998; Kidd, 2011; Reber, 1967; Rebuschat, 2013). Studies on statistical learning have shown that implicit learning mechanisms facilitate children's ability to infer word boundaries (e.g., Aslin et al., 1998; Gómez, 2002; Gómez & Gerken, 1999; Saffran et al., 1996), acquire words and their meanings (e.g., Smith et al., 2014; Smith & Yu, 2008; Vlach & Johnson, 2013), and form grammatical categories (e.g., Lany, 2014; Lany & Saffran, 2013). Moreover, several studies show a relationship between statistical learning capacities and language skills in clinical populations (for developmental language disorders see for example Evans, et al., 2009 and Lammertink, et al. 2017; for

---

* This chapter is a slightly modified version of a manuscript that is ready to be submitted: Spit, S., Andringa, S., Rispens, J., & Aboh, E.O. (under review at *Language Development Research*). The effects of instruction, input and sleep during kindergartener's acquisition of grammatical structures: an eye tracking study.

developmental dyslexia see for example Gabay et al., 2015), although the strength and extent of this relationship is under debate (Lammertink, Boersma, Wijnen & Rispens, 2020; van Wittenloostuijn, et al., 2017).

Implicitly detecting statistical regularities thus seems to contribute to language acquisition in both typical and atypical learners. Nevertheless, it is unlikely that this process alone can explain how languages are acquired (Lidz & Gagliardi, 2015; Yang & Montrul, 2017). Studies on adults indicate that language acquisition can be influenced by, for example, explicit instruction. Explicit instruction is often defined as a type of instruction in which attention is focused on the mapping between form and function in the linguistic input (DeKeyser 1995; R. Ellis, 2005, 2009; Housen & Pierrard, 2006). Explicit instruction can thus make learners aware of the regularities they need to acquire (e.g. Hamrick & Rebuschat, 2012), and such awareness can be beneficial when learning linguistic structures, especially in instructed contexts. Regardless of the linguistic domain (such as pragmatics, syntax, morpho-syntax, phonetics), adults often show higher learning gains when they are made aware of the target structure via explicit instruction compared to uninstructed language learning peers (Goo et al., 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). The role of awareness and explicit instruction in younger learners has been investigated as well, but seems more restricted to a limited set of linguistic domains. Meta-analyses indicate explicit instruction can be beneficial when kindergarteners acquire vocabulary items (Marulis & Neuman, 2010), or when preschoolers, kindergarteners, and first graders have to learn how to read and spell (Ehri et al., 2011). Lichtman (2016), however, rightly observed that relatively little is known about the role of explicit instruction when young learners acquire structures of a more grammatical nature. She therefore investigated the effect of explicit instruction in a group of children from 5 to 7 years old, but did not find an effect of such instruction on learning word order and a determiner system. Crucially, this study did not show an effect in the adult control group either, which makes it difficult to determine whether explicit instruction had no effect on learning in this particular age group. Although the results from this experiment might be difficult to interpret, it highlights the importance of further investigating the role of explicit instruction in younger learners.

In Chapter 5, we therefore tried to gain more insight into this issue by exposing 4 to 7-year-old kindergarteners to a miniature language from which they had to learn a number marker. The meaning of this marker, which indicated whether nouns in the sentence should be interpreted as plural or not, could only be learned from the distributional properties in the input. To learn this regularity, children were exposed to the language for approximately 20 minutes. Importantly, half of the children received only this exposure, whereas the other half was also explicitly instructed about the presence and form of the number marker. Afterwards, all children were tested on their knowledge of the number marker using a picture matching task during which we measured their eye movements. Measures like grammaticality judgment tasks, but also the picture matching task in our experiment, have been suggested to be more strongly associated with explicit knowledge, whereas eye tracking is assumed to tap more into implicit knowledge (Andringa & Rebuschat, 2015; Bowles, 2011; R. Ellis, 2005; Godfroid et al., 2015; Han & Ellis, 1998).

In this distinction, explicit knowledge refers to knowledge that is available to awareness, whereas implicit knowledge is not (Rebuschat, 2013). A picture matching task focusses on the form and grammaticality of linguistic stimuli rather explicitly, because participants are directly asked to match a linguistic form to a particular visual interpretation. Using this type of task therefore makes it more likely that explicit knowledge drives behavior of the participant (R. Ellis, 2005; Loewen, 2009). Measuring eye movements, however, does not involve asking participants anything explicitly about the grammaticality of test items, and thus makes it less likely that explicit knowledge drives task performance (R. Ellis, 2005; Sagarra & Seibert Hanson, 2011). Because explicit instruction can make learners aware of the target structure, and as a result might affect the development of implicit and explicit knowledge in different ways (Godfroid, 2015), it was important to include these two different measures (e.g. Andringa et al., 2011).

The results in Chapter 5 showed that explicit instruction did not result in higher accuracy rates in the picture matching task measure, but that it did lead to earlier predictive eye movements during the same task. This might indicate that explicit instruction has an effect on young learners acquiring a grammatical marker, but that this effect differs depending on how knowledge is measured. The results could suggest that different types of knowledge are influenced differently by explicit instruction: explicit instruction influences more implicit knowledge in this age groups, as reflected in their eye movements, whereas it affects explicit knowledge less so, as reflected by their performance in the picture matching task. Alternatively, the effect of instruction we observed in the eye movements could be a precursor of better performance in the picture matching task, but that learning needs more time to be established. This would mean that the children were tested too soon for the effect of explicit instruction to appear in the accuracy measure. Perhaps the impact of explicit instruction that we observed in children's eye movements only becomes visible in accuracy data after a longer consolidation period, or when children receive more input. In this paper, we therefore present an exact replication of the previous study, but included tests on a second day as well. On the first day of this experiment, we expect to find similar results to those we observed in our previous study. If we find comparable or bigger learning effects on the second day, this could show that what happens on day one is indeed a precursor of more robust learning. Such learning perhaps only occurs after a consolidation period including sleep, or by receiving more input to learn the grammatical marker. Adding a second day of testing to our experiment allows us to investigate precisely these factors.

Consolidation is defined as the process during which information that is stored in our memory gets strengthened and enhanced (Axelsson, Williams & Horst, 2016; Dewar, Alber, Cowan & Della Sala, 2014; Dudai, 2002). After information is encountered, it needs to be encoded immediately and stored in our long-term memory (Dudai, 2004). Several factors influence these reorganization processes, and thus play a role in *not* forgetting what one has experienced. Mental representations are simply consolidated when a learner receives more input, and encoded information becomes represented more strongly (Ellis, 2002; Bybee, 2010). Furthermore, both wakeful rest and sleeping time after encountering a stimulus can be beneficial for strengthening a stored representation (Dewar, Alber, Butler, Cowan & Della Sala, 2012; Mednick, Cai, Shuman, Anagnostaras & Wixted, 2011; Wixted, 2004 for wakeful rest; Diekelmann, Wilhelm & Born, 2009; Stickgold & Walker, 2013 for sleep): without the presence of new sensory stimuli, encoded representations are enhanced during such intervals.

There is cumulative evidence that memory consolidation plays an important role when adults acquire linguistic structures using implicit learning mechanisms (Feld & Diekelmann, 2015). When adults need to learn grammatical structures implicitly, they generally perform better on a delayed post-test than an immediate post-test (Spada & Tomita, 2010), which could be interpreted as a consolidation effect. Some studies even suggest such effects are stronger for sleep-related consolidation than for wakeful consolidation, although effects are found for both types of rest (Frost & Monaghan, 2017). Similar to what is observed in adults, memory consolidation also plays an important role when children have to process and store information (Diekelmann et al., 2009; Fisher, Wilhelm & Born, 2007). Wilhelm, Diekelmann & Born (2008), for example, show that 6 to 8-year-old children who have to match pairs of words (i.e. *arm* & *blood*), are better at recalling those words after an interval of sleep, than after staying awake for the same period of time. Furthermore, Van den Berghe (2019) presents evidence for consolidation in vocabulary acquisition: a series of experiments shows that kindergarteners who acquired words in their L2 showed better knowledge during a delayed post-test than during an immediate post-test. Sleep related memory consolidation likely also influences the acquisition of more grammatical patterns, as Gómez et al., (2006) showed in an artificial grammar learning experiment. Infants who napped after being exposed to a novel stream of sounds, were able to pick up the co-occurrence of particular syllables in this stream, whereas children that did not nap were not able to pick up these regularities.

The goal of the present study was therefore to replicate the study from Chapter 5, as well as to investigate the effect of consolidation. Results from our earlier study showed an effect of explicit instruction on acquiring a number marker that was difficult to explain: children who received explicit instruction did not show higher accuracy rates during a picture matching task measuring knowledge, but did show earlier predictive eye movements during the same task. We used the artificial language from this previous study, which was also used in Chapters 2 and 4, and allowed for an exact replication of this prior study (see Marsden et al., 2018 for why we need more replication work in linguistics). Instead of exposing and testing children only on the same day as in the previous study, children were tested one day later as well. On this second day, half of all children received extra input to acquire the grammatical marker. This design enabled us to investigate whether a consolidation period including sleep, and/or prolonged input help kindergarteners when they acquire a morphosyntactic element, and also whether such factors interact with explicit instruction. If the results on the second day are similar to those on the first day, this would support the idea that the explicit instruction affects the development of implicit and explicit knowledge of the marker differently. However, we might also find that learning effects in the picture matching task increase on the second day of testing. This would mean that children only benefit from explicit instruction when they are given the opportunity to consolidate this knowledge. We might also see that a difference in accuracy results on day 1 and day 2 depends on the amount of input children receive. Such a finding would indicate that explicit instruction or consolidation only influences learning when children receive a certain amount of exposure.

6.1 Method

6.1.1 Participants

103 Dutch speaking children (50 males, 53 females, *M* = 5;4 years, *SD* = 0;8) took part in this experiment. All children were in kindergarten and were recruited from 7 primary schools in the western and southern regions of the Netherlands, and in Flanders. They did not have any diagnosed language or communication disorders based on teacher reports. No restrictions were imposed upon taking part in the experiment, and no further background information was collected. 5 participants had to be excluded on the first day of testing, because of problems with calibrating (*N* = 3) or other technical issues (*N* = 2). As a result, we present data from 98 Dutch speaking children (49 males, 49 females, *M* = 5;4 years, *SD* = 0;8) on the first day of testing. From this group, 52 children received explicit instruction during the training (26 males, 26 females, *M* = 5;5 years, *SD* = 0;7), while the other 46 children did not receive such instruction and were only exposed to the miniature language (23 males, 23 females, *M* = 5;3 years, *SD* = 0;8). In the remainder of this chapter we refer to the former group as the explicitly instructed group and to the latter as learners who received only exposure.

From the 98 children who participated on day one, 17 children could not participate on the second day, because of technical issues (*N* = 8), illness (*N* = 4) or other reasons (*N* = 5). We thus present results from 81 children on the second day. From this group, 42 children received extra input on the second day

(21 males, 21 females, *M* = 5;5 years, *SD* = 0;8), the other 39 children only participated in another test phase on the second day (20 males, 19 females, *M* = 5;4 years, *SD* = 0;8). Children tested on the second day can thus be divided in four groups: children who received explicit instruction and extra input (11 males, 11 females, *M* = 5;6 years, *SD* = 0;7), children who received only exposure and extra input (10 males, 10 females, *M* = 5;1 years, *SD* = 0;9), children who received explicit instruction, but no extra input (10 males, 9 females, *M* = 5;5 years, *SD* = 0;8), and children who received exposure only and no extra input (10 males, 10 females, *M* = 5;4 years, *SD* = 0;8).

6.1.2 Materials

We used the artificial language from Chapters 2, 4 and 5, which consisted of four proper names, three verbs, two grammatical markers, six frequent nouns, twelve infrequent nouns and one conjunction. Apart from the proper names, all words were novel. The language had subject-verb-object word order. Recall from the discussion in previous chapters that, in this language, a noun phrase on its own does not encode number and could correspond to both singular and plural referents. In a sentence, however, an argument noun phrase must be introduced by a nominal marker (*pli* or *tra*). In our miniature language, a noun introduced by *tra* can refer to both singular and plural referents: the correct interpretation must be inferred from the visual context. The language includes another nominal grammatical marker *pli*, which encodes number and indicates that the noun necessarily refers to multiple referents. In short, the rule participants had to learn was that whenever *pli* preceded a noun, this noun always referred to multiple referents.

6.1.3 Structure of the experiment on day 1

The overall structure of the experiment on day one was exactly the same as the experiment described in Chapter 5, but will be shortly described again below. To learn the regularity, participants were asked to help four protagonists learn a new language. They were told they would see pictures and hear things in the new language that matched the pictures they saw. After this introduction, the experiment consisted of four parts. It started out with a small vocabulary training. After the vocabulary training a rule training session followed. During this session, children received input to acquire the grammatical rule. Children who were in the explicit instruction group received instruction immediately before the second rule learning phase, after they were exposed to 1/3 of the training sentences. During this instruction they were told the meaning of both grammatical markers *pli* and *tra* by one of the characters in the game. The Dutch instruction and its English translation can be found in Chapter 5.

After the rule training, participants performed a picture matching task to test whether they acquired the rule. In the picture matching task, participants heard sentences from the rule training phase, and had to choose which of two pictures matched each sentence. For the test items of this task, participants had to choose between pictures with either one or multiple referents. The two pictures always depicted different referents. 12 of the experimental items contained *pli*. For these items, the target picture always showed a referent depicted more than once. The other 12 of the experimental items contained *tra*. Any number of referents would be grammatical for these items. Sensitivity to the statistical regularity in the input would lead to more target answers on trials with *pli* than on trials with *tra*, because for trials with *pli* participants could base their response on the number of referents the picture showed, whereas the number of referents was not indicative for trials with *tra*. To test whether explicit instruction also influenced more implicit knowledge, we measured participants' eye movements during the picture matching task. If participants acquired the regularity, this would lead towards more correct looks towards the target after the onset of the grammatical marker for sentences with *pli,* than for sentences with *tra*.

6.1.4 Experiment on day 2

To investigate whether not only a consolidation period, but also extra input influences acquisition of the grammatical marker, children who were able to participate again in the experiment on the second day of testing were divided into two groups. Half of the children received additional input to learn the grammatical regularity, while the other half of the children did not receive additional input. Children who received additional input participated in an additional training phase on the second day. This training phase contained the exact same input to acquire the number marker as the rule training phase on the first day, but without a vocabulary training and test before the first block of training sentences. After this additional rule training phase, children performed a picture matching task. Children who did not receive more input skipped the rule training phase, and started with a picture matching task. For all children, the picture matching task on the second day of testing was identical to the picture matching task on the first day of testing. After the picture matching task on the second day, an informal debriefing took place for all children, in which we examined whether participants were aware of their knowledge in a way they could verbalize. During this debriefing, participants were asked how they knew what was the correct answer during the picture matching task, and whether they knew the meaning of the words *pli* and *tra*. None of the children received any further explicit instruction about the grammatical rule on the second day of the experiment. Figure 6.1 provides a visual illustration of the structure of the experiment on this second day.

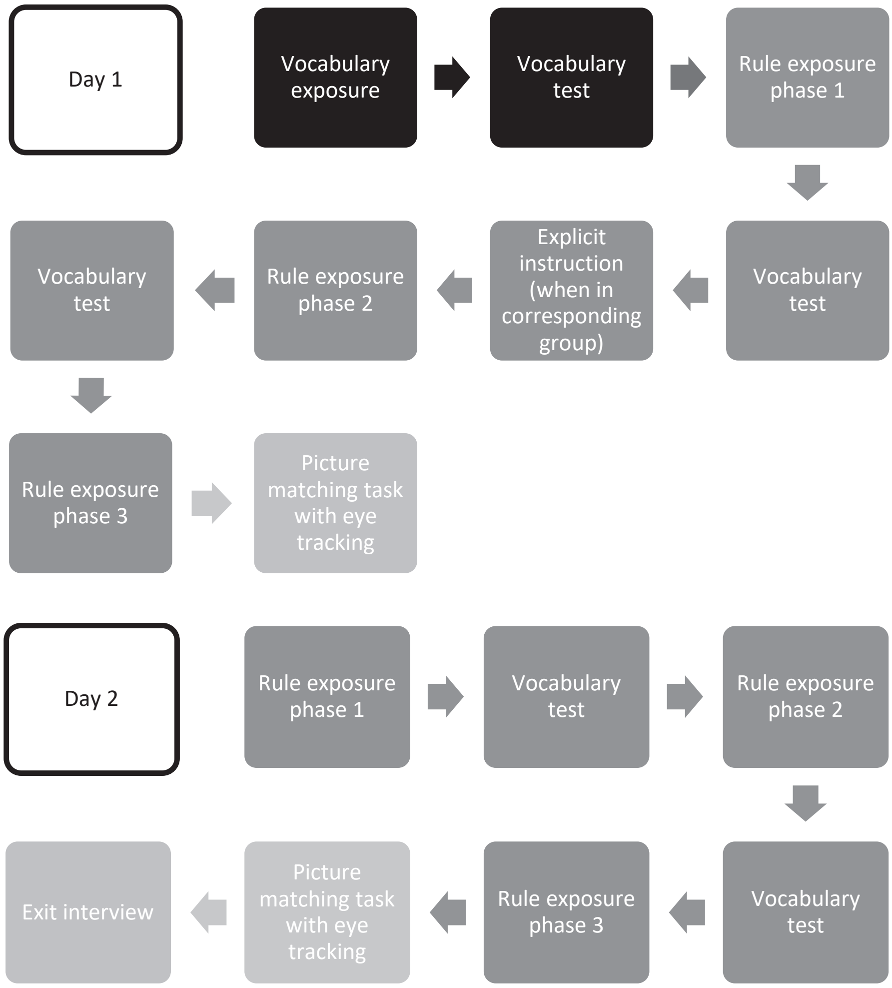


Figure 6.1. A visual representation of the structure of the experiment on the two days. The top boxes on day one show the parts of the vocabulary training. The lighter boxes show the different parts of the rule training. The lightest boxes show the test phase. Only children who received extra input took part in the three exposure phases on day two.

It is important to note that, apart from the extra activities on the second day of testing, the current experiment differs in one other aspect from the experiment from Chapter 5. In the previous experiment, children took part in debriefing on the first day of testing, since this was the only day they participated in the experiment. In the current experiment, this debriefing took place on the second day of testing. During this debriefing, participants were asked how they knew what the correct answer during the picture matching task was, and whether they knew the meaning of the words *pli* and *tra*. Besides the debriefing, the experiment as conducted on day one was exactly identical to the study reported in Chapter 5. All minor changes between these two eye tracking studies and the other two studies from Chapter 2 and 4, are discussed in Chapter 5 as well.

6.1.5 Procedure

Stimuli were the same as those used in Chapters 2, 4 and 5 and were recorded by a female native speaker of Dutch. Sentences were recorded as a whole, leading to minor natural variation between similar words in the language. The test was administered in a quiet room at the participants' school. The task was presented on a laptop using E-prime (Psychology Software Tools, 2012). Eye tracking data was gathered using a Tobii Pro X3-120 Eye Tracker with an External Processing Unit. During vocabulary training, nouns and their accompanying pictures were presented for three seconds, before automatically moving to the next noun. During rule training, sentences and their accompanying pictures were presented for four seconds, before automatically moving to the next sentence. The instruction was prerecorded and segmented into smaller bits, such that the instructor could press a button on the keyboard to continue with the next piece of instruction. During the test phase, the experimenter pressed a button on the keyboard that corresponded to the answer the participant gave. Scores from this phase were registered automatically.

We calibrated the eye tracker before the start of the test phase, and sampled eye positions at a rate of 120 Hz. After each test item, participants saw a fixation cross in the middle of the screen. The following test item started if participants fixated for 100 ms (12 samples) on the cross. This ensured that participants were looking at the middle of the screen, between the pictures, at the onset of every test item. For every test item, the audio and pictures were presented simultaneously. Both the participant and the experimenter listened to the audio using headphones. The vocabulary training lasted 8 minutes on average, the rule training phase 15 minutes, and the picture matching task including calibration 10 minutes. The full experiment took approximately 30 minutes per participant on the first day. On the second day, the experiment lasted approximately 10 or 25 minutes, depending on whether children participated only in the picture matching task, or received extra input during another rule training phase. Ethical approval for this study was obtained from the University of Amsterdam and active consent was obtained from children's parents or legal guardians before the start of the study.

6.1.6 Analysis

We used linear mixed effects modeling to analyze the results of the picture matching task and cluster based permutation analyses to analyze the eye tracking data. All analyses were carried out in R (R Core Team, 2015) using the lme4 package (Bates et al., 2015) and eyetrackingR package (Dink & Ferguson, 2015) where needed. To see whether we could replicate the results from Chapter 5, we first used linear mixed effects models and cluster based permutation analyses to investigate the results of the first day of testing only, using data from all our 98 participants. Additional analyses were run to investigate the effects of consolidation and extra input. These analyses were based on data from the 81 children who participated on both days.

To investigate the effect of explicit instruction on learning as measured by the picture matching task on day one, we carried out a generalized linear regression model with mixed effects that was similar to the model we reported on in Chapter 5. This model took the responses from the task (1 for a correct answer, 0 for an incorrect answer) as a dependent variable, marker type (*pli* vs. *tra*) as a within-participants fixed effect, type of instruction (explicit instruction vs. exposure only) as a between-participants fixed effect, participant as a between-participants random effect and item as a within-participants random effect. Our fixed effects were included in this model, because we were *a priori* interested in their contribution to the outcome (Gelman & Hill, 2007). Orthogonal sum-to-zero contrast coding was applied to our binary fixed effects (i.e., marker type and instruction type; Baguley, 2012, p590-621). We aimed to keep the models as fully specified as possible by including random intercepts for participants and items (Barr et al., 2013). We increased the number of iterations to 100.000 (Powell, 2009), to solve potential issues with non-converging models. This enabled us to report on a random effect structure justified by our data (Jaeger, 2009). We report simple rather than standardized effect sizes (Baguley, 2009) and Wald confidence intervals (Agresti & Coull, 1998).

To investigate whether learning was not only influenced by explicit instruction, but also by consolidation and receiving additional input on learning, we carried out another generalized linear regression model with mixed effects to analyze scores from both days on the picture matching task. We used a similar procedure as described above to determine which model we should report on. The model we report on took the responses from the task (1 for a correct answer, 0 for an incorrect answer) as a dependent variable. Marker type, and day of testing (day one vs. day two) were included as within-participants fixed effects, type of instruction and input condition (extra input vs. no extra input) as between-participants fixed effects, as well as the possible interactions between all four fixed effects. The model took participant as a between-participants random effect and item as a within-participants random effect. No random slopes were included in the model because it would not converge with these included.

To investigate looking behavior during the picture matching task, we carried out several cluster based permutation analyses (see Curcic et al., 2019; Dink & Ferguson, 2015; Maris & Oostenveld 2007) using the eyetrackingR package (Dink & Ferguson 2015). This type of analysis enabled us to investigate the development of the proportion of looks towards the target picture over time. We will describe the general procedure that we followed for this analysis, which was similar for the analyses that focused on day one only, and for the analyses that compared both days of testing.

We analyzed trials from 500 ms before the onset of the grammatical marker until 4000 ms after the onset of the marker (see Figures 6.4-6.10 for a visual illustration of this time window). On average a grammatical marker lasted 350 ms, there was approximately 200 ms between the offset of the marker and the onset of the noun, and a noun lasted about 550 ms. A sentence thus lasted for around 1100 ms from the onset of the marker. This means the endpoint of the frame we analyzed was approximately 3000 ms after the offset of the sentence. As children often took at least 3000 ms to give an answer during the task, we have eye tracking data for the majority of the test items in this timeframe. After this point in time, we have fewer data points available, as children might have given an answer, and the next test item would be played. When interpreting the eye tracking data, readers should take into account that in this age group it takes approximately 400 ms before eye movements reflect the processing of a presented stimulus (Clackson et al., 2011; Fukushima et al., 2000).

For the cluster-based permutation analyses, we analyzed all test items, regardless of whether participants gave a correct or incorrect answer during the picture matching task. For every sampled look we coded whether it was directed at the target picture (1) or not (0). The area of interest for both types of pictures was the entire area within the colored frames. We removed all samples that were irrelevant, i.e., looks on screen but not on either of the two pictures, or missing. After this, we excluded trials in which more than 50% of the samples had been removed as a result of this procedure. This analysis, which involves two steps, allowed us to determine at which parts within the marked time frame participants were likely to look more often towards the target picture for sentences with *pli* than for sentences with *tra*. In the first step, the data were split into 50 ms (6 samples) time bins, and by calculating a *t*-value for each time bin we determined whether participants looked significantly more often towards the target picture for *pli* sentences than for *tra* sentences. Adjacent time bins for which there are significant effects were clustered together and the *t*-values were summed. In a second step, the data were reshuffled hundreds of times, to see how likely the observed clusters and accompanying summed statistics would be if the data were randomly generated. Using this procedure, we can correct for false alarms, while not losing much of the sensitivity of the data. The expectation was that children who acquired the plurality marker would be more likely to look at the correct image in *pli* sentences than in *tra* sentences.

A cluster based permutation analysis allows for one predictor only, making it impossible to look at interactions between marker and instruction type, day of testing and input condition simultaneously. Therefore, we had to run several analyses to compare the different experimental conditions. We first carried out analyses that were similar to the ones we report on in Chapter 5, to allow for comparison between these studies. For these analyses we used data from all 98 participants of the first day of testing. First we looked at all learners together, to see whether they looked more often towards the target picture for sentences with *pli* than sentences with *tra*. Next, the results from the explicitly instructed group and the uninstructed group were analyzed separately.

We carried out another set cluster-based permutation analyses to investigate the data from the 81 children that participated on both days. Since we could not include day of testing as a predictor in these analyses, we had to run separate analyses for two days for each group comparison that we wanted to make. For example, when looking at all learners together, we ran one analysis for all learners on day one to see whether they looked more often towards the target picture for sentences with *pli* than sentences with *tra*, and another analysis for all learners on day two to investigate the same looking behavior. We report on similar analyses in which we compare the two days for explicitly instructed and uninstructed learners, regardless of their input condition, and for children who received extra input on day two and children who did not receive extra input on day two, regardless of their instruction condition.[3]

Before discussing our results, we should make clear what kind of results we might expect. In general, we expect that more correct answers for sentences with *pli* than for sentences with *tra* during the picture matching task, and that more looks towards the target for the former type of sentences than for the latter, reflecting some form of learning of the grammatical marker. For the analyses that only looked at day one of testing, we expected to find similar results to the study from Chapter 5. Children seemingly showed learning in the picture matching task, and instruction had no positive impact on this. All children showed the looking behavior that we described, but instructed learners showed this behavior earlier during the test trials than uninstructed learners. For the analyses that compare the two days, it is more difficult to formulate clear expectations, but we

[3] We also carried out similar analyses comparing the two days of testing for the smaller subgroups: explicitly instructed learners who received extra input on second day of testing; explicitly instructed learners who did not receive extra input on second day of testing; uninstructed learners who received extra input on second day of testing; uninstructed learners who did not receive extra input on second day of testing. Because these subgroups were relatively small, and seemingly lacked sufficient statistical power, and because of limits of space, we will not report on the latter analyses here. The analyses that investigate the subgroups can be consulted on our OSF page.

described some scenarios earlier. Possible outcomes could be that results on the second day remained similar to those on the first day, which would support the idea that the explicit instruction affects the development of implicit and explicit knowledge of the marker differently. In this case, we would observe a main effect of marker in the picture task, but no interactions between day of testing and learning the marker, and should see similar eye tracking behavior on both days of testing. Alternatively, we might find a positive interaction between day of testing, instruction and grammatical marker, which would show children only benefit from explicit instruction when they are given the opportunity to consolidate this knowledge. The eye movements of these children would then remain similar across the two days. We might also see that a difference in learning between days depends on the amount of input children receive, which would be supported by an interaction between day of testing, input group and grammatical marker. In this case, we might expect children who receive additional input to show the expected looking behavior in earlier time frames or in time windows that last longer on the second day of testing.

6.2 Results

6.2.1 Picture matching task: replication

An overview of the descriptive statistics from the picture matching task of day one from all 98 children who participated on that day can be found in Table 6.1 and Figure 6.2. We first present the results from all children to see whether the results as obtained on day one were in line with the previous study. For this group, we wanted to know whether they would give more correct answers for sentences with *pli* than sentences with *tra*, which would reflect learning of the grammatical marker. We also considered whether this behavior interacted with the type of instruction children received. Importantly, the observed effects exactly mirrored those observed in Chapter 5, both in direction and size. The table shows that participants gave more target answers when sentences contained *pli* than when sentences contained *tra*, but this effect was not significant (OR = 1.227, 95% CI = [0.961, 1.565], z = 1.642, p = .101). Results also show that explicitly instructed learners gave slightly less target answers in general than learners who received exposure only, but this effect was not significant either (OR = 0.916, 95% CI = [0.778, 1.079], z = -1.044, p = .296). Additionally, we observed a non-significant interaction between type of instruction and grammatical marker (OR = 0.939, 95% CI = [0.677, 1.303], z = -0.375, p = .708). Although explicitly instructed learners showed less learning than their peers who received only exposure, this effect was negligible and not significant.

Table 6.1. Scores from the picture matching task.

	Explicit Instruction (*N* = 52)			**Exposure only (*N* = 46)**		
	M	**SD**	**Range**	**M**	**SD**	**Range**
Pli	6.27	1.76	2-10	6.57	1.95	2-10
Tra	5.77	1.71	2-11	5.91	1.47	3-10

Note: Scores indicate the number of target answers produced by children who received explicit instruction and exposure, and children who received only exposure on day one, for all 98 children who participated in the experiment. Sentences with grammatical marker *pli* were predictable, which should lead to more correct target answers in the case of learning. Scores could range from 0-12.

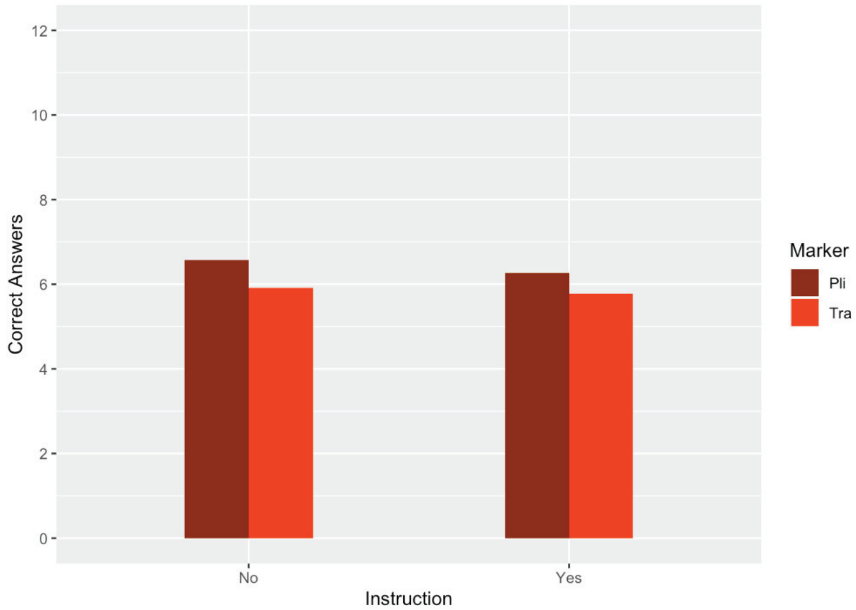


Figure 6.2. A graph depicting the results from the picture matching task for uninstructed (left) and explicitly instructed (right) learners of the first day for all 98 participants. Scores from the picture matching task indicate the number of target answers. Scores could range from 0-12.

6.2.2 Picture matching task: comparing days

To investigate how a consolidation period including sleep and extra input influenced learning the marker, we carried out separate analyses that looked at the results on the picture matching task from the 81 children who participated on both days. The descriptives for this group can be found in Table 6.2 and Figure 6.3. We wanted to know whether the day of testing would interact with learning. Across both days, children produced more target answers when sentences contained *pli* than when sentences contained *tra*, but this effect was not significant (OR = 1.190, 95% CI = [0.951, 1.488], z = 1.523, p = .128). We also observed that participants gave more target answers for both sentence types combined on the second day of testing (*OR* = 1.170, 95% CI = [1.029, 1.330], z = 2.404, p = .016). Yet, we did not find a significant interaction between marker type and the day of testing (*OR* = 0.918, 95% CI = [0.710, 1.186], z = -0.656, p = .512), which would have been indicative of higher sensitivity to the regularity on the second day of testing than on the first day of testing. Results also show a marginally significant interaction between day of testing and instruction condition (*OR* = 1.287, 95% CI = [0.998, 1.663], z = 1.930, p = .054), which seems to show that explicitly instructed learners gave slightly more target answers in general on the second day than learners who received exposure only. However, we did not observe a significant interaction between type of instruction, grammatical marker and the day of testing (*OR* = 0.870, 95% CI = [0.521, 1.452], z = -0.532, p = .595), which would have been indicative of an instruction effect that is different depending on the day of testing. We did observe a significant interaction between input condition, grammatical marker and day of testing (*OR* = 1.678, 95% CI = [1.006, 2.801], z = 1.982, p = .048), which suggests that the extra input on the second day of testing had an effect on picture matching task performance on the second day of testing. Children who did not receive additional input showed a slight decrease in their scores on the second day, whereas children who received extra input showed a slight increase.

Table 6.2. Scores from the picture matching of all children who participated on both days.

		Day 1			Day 2		
		M	**SD**	**Range**	**M**	**SD**	**Range**
All learners (*N* = 81)	*Pli*	6.40	1.91	2-10	6.77	2.06	2-12
	Tra	5.78	1.68	2-11	6.37	1.63	3-10
EI learners (*N* = 41)	*Pli*	6.37	1.89	2-10	6.95	1.90	2-11
	Tra	5.73	1.84	2-11	6.76	1.56	4-10
EO learners (*N* = 40)	*Pli*	6.42	1.95	2-10	6.58	2.22	2-12
	Tra	5.83	1.52	3-10	5.97	1.62	3-9
AI learners (*N* = 42)	*Pli*	6.21	1.99	2-10	7.21	2.10	2-11
	Tra	5.83	1.94	2-11	6.33	1.71	3-10
No AI learners (*N* =39)	*Pli*	6.59	1.82	3-10	6.28	1.92	2-12
	Tra	5.72	1.38	3-9	6.41	1.57	3-9

Note: Scores indicate the number of target answers produced. Scores are first shown from all learners, then split by instruction type (EI = Explicitly Instructed, EO = Exposure Only), then by input condition (AI = Additional Input on day two, No AI = No Additional Input on day two). For a further division in subgroups see our OSF page. Sentences with grammatical marker *pli* were predictable, which should lead to more correct target answers in the case of learning. Scores could range from 0-12.

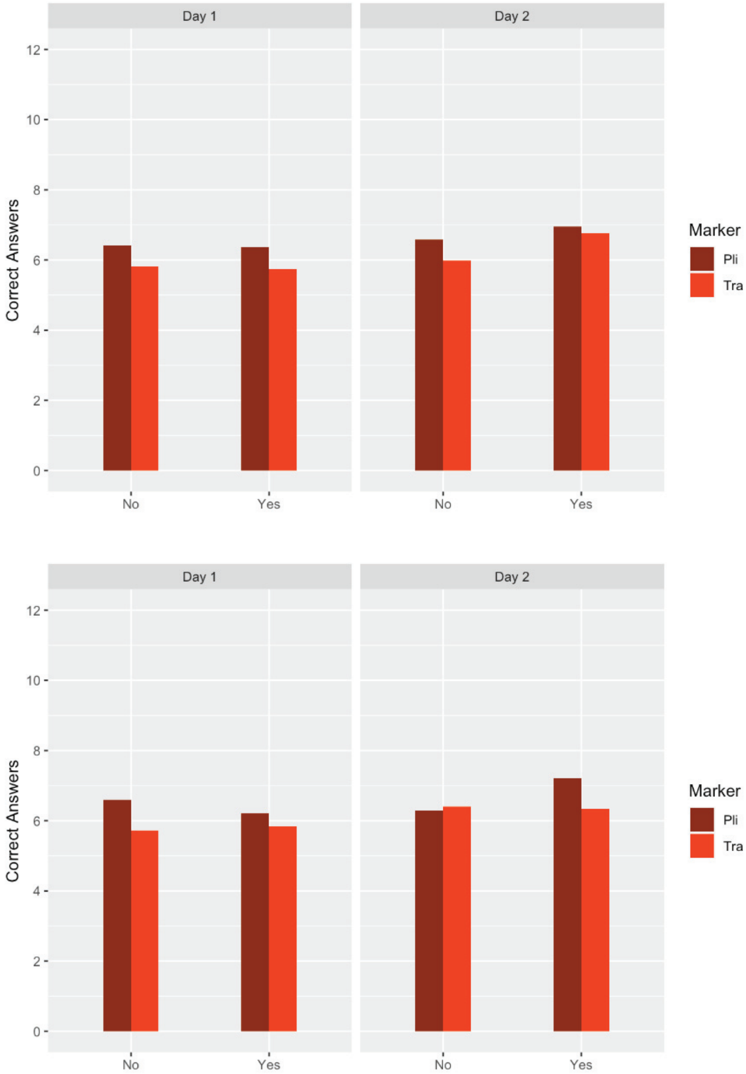
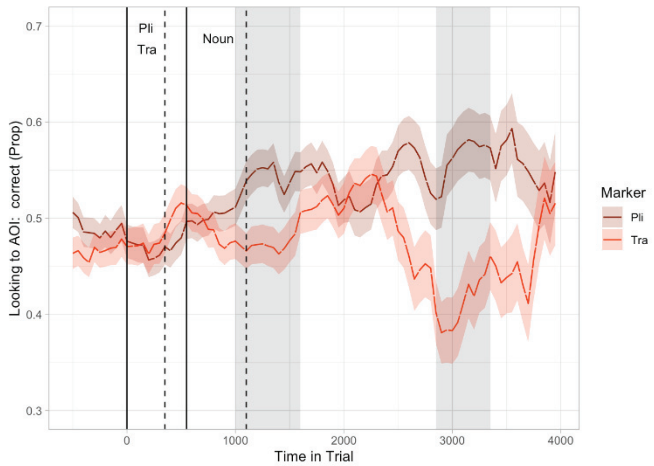


Figure 6.3. Two graphs depicting the results from the picture matching task for all 81 children who participated on both days. The top graph shows the results split by instruction type (Yes = Explicit instruction, No = Exposure only). The bottom graph shows the results split by input condition (Yes = Additional input on day two, No = No additional input on day two). Scores from the picture matching task indicated the number of target answers. Scores could range from 0-12.

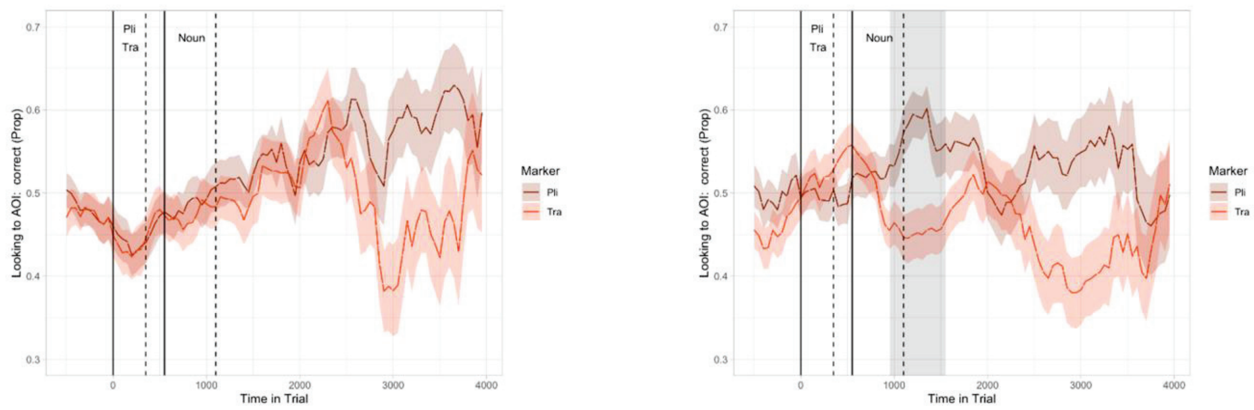
6.2.3 Eye tracking data: Replication

As a first step in the analysis of the eye tracking data, we looked again at data from all 98 children who only participated on the first day of testing, and wanted to see whether we found similar looking behavior as in our previous experiment. To test this, we conducted a cluster based permutation analysis with the proportion of looks towards the target picture as a dependent variable, and grammatical marker as a within-participants factor. We carried out this analysis first for all explicitly instructed learners and learners who received only exposure together. For all learners combined, the first step of the analysis showed that participants behaved differently for sentences with *pli* than for sentences with *tra* during five time clusters. In these cases, positive sum statistics were observed indicating that participants were more likely to look towards the target picture in *pli* sentences than in *tra* sentences. This is the looking behavior we expect children to show, if they have learned the rule. Figure 6.4 shows that this effected turned out to be significant for two clusters: one that started shortly after the onset of the noun, and one that occurred well after the end of the sentence. For instructed learners we found one cluster during which participants showed the expected looking behavior, but this effect was not significant (Figure 6.5, left panel). For exposure only participants, four clusters were identified (Figure 6.5, right panel), of which only the second was found significant in step 2. Importantly, in the groups as a whole, and in the uninstructed learners, we found the expected looking patterns, during time windows that are comparable to those we observed in Chapter 5. Although instructed learners showed early predictive eye movements in our previous study, we were unable to observe any predictive eye movements in these learners in the present study.



Cluster	Sum statistic	Time range (ms)	*p*-value
1	**27.09**	**1000 – 1600**	**.038**
2	9.23	2550 – 2750	.280
3	**29.39**	**2850 – 3350**	**.030**
4	7.00	3450 – 3600	.352
5	4.61	3650 – 3750	.471

Figure 6.4. All learners combined, day one: on the top, proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra* for all 98 children who participated on day one. On the bottom, results of the cluster based permutation analysis for the same group. Proportion of looks was calculated from all looks that were either on the target or on the non-target picture. Solid lines indicate the onset of the grammatical marker (*pli* or *tra*) and the noun respectively, dashed lines the offsets of these words. Time range is measured from the onset of the grammatical marker. The shaded area indicates the cluster of time bins that turned out significant after the second step of the analysis and for which an effect of grammatical marker was found. This cluster is marked in bold on the right.



Day 1

Cluster	Sum statistic	Time range (ms)	*p*-value
1	11.60	2850 – 3100	.211

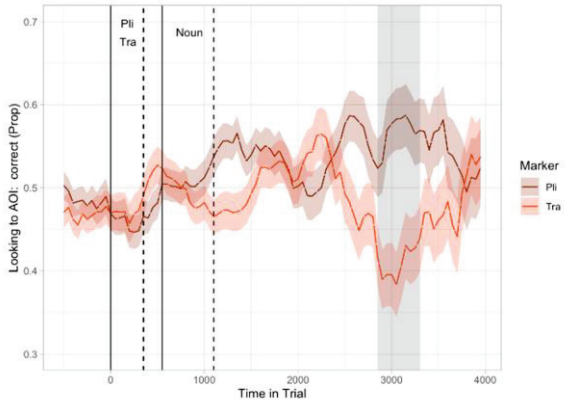
Day 2

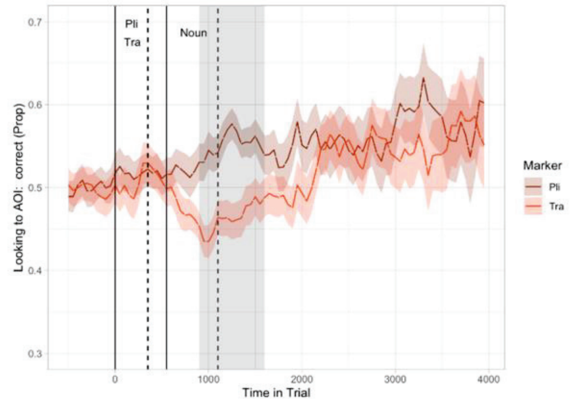
Cluster	Sum statistic	Time range (ms)	*p*-value
1	-4.37	450 – 550	.557
2	**36.03**	**950 – 1550**	**.014**
3	18.43	2950 – 3350	.113
4	2.07	3450 – 3500	.725

Figure 6.5. Instructed (left) and uninstructed learners (right), day one: on the top, proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra* for all instructed and uninstructed children who participated on day one. On the bottom, results of the cluster based permutation analyses for the same groups. The left picture and table show the results from day one, the right picture and table those from day two. The cluster in bold turned out significant after the second step of the analysis.

6.2.4 Eye tracking data: Comparing days

Next, we analyzed data from only the 81 children who participated on both days of testing. Here, we were also interested to see whether looking behavior changed on the second day as a consequence of consolidation or because of the extra input provided on the second day of testing. The left panel of Figure 6.6 shows three clusters during which participants looked more often towards the target picture for sentences with *pli* than for sentences with *tra*. Of these clusters, only the one from 2850 to 3300 ms reached significance. The cluster from 1100 to 1450 was not significant, while it was in the entire group of 98 (see Figure 6.4), which is most likely to be due to the reduced sample size and resulting decreased statistical power. On the second day we observed that the whole group looked significantly more often towards the target picture for sentences with *pli* than for sentences with *tra* in a relatively long time cluster. This cluster is comparable to the first of three clusters that we saw on day one.





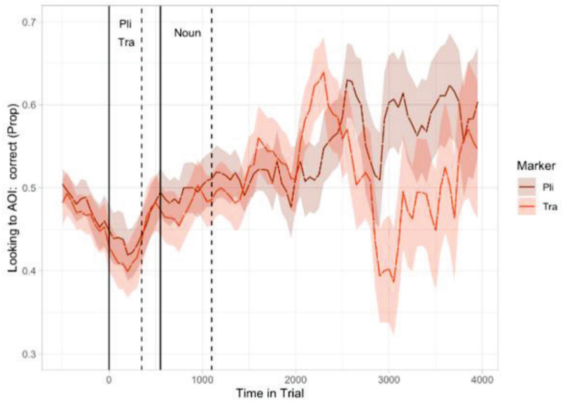
Day 1

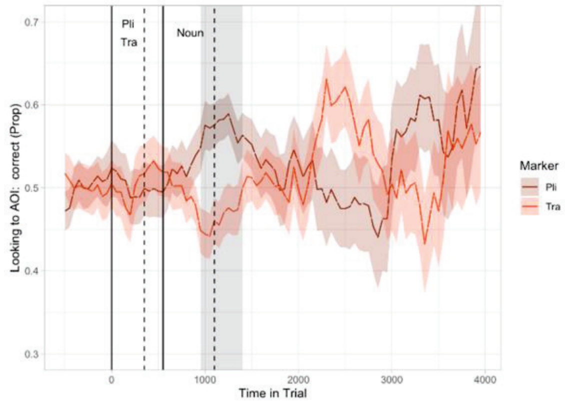
Cluster	Sum statistic	Time range (ms)	*p*-value
1	17.85	1100 – 1450	.115
2	6.68	2550 – 2700	.397
3	**26.03**	**2850 – 3300**	**.048**

Day 2

Cluster	Sum statistic	Time range (ms)	*p*-value
1	**43.32**	**900 – 1600**	**.003**

Figure 6.6. All learners, both days: on the top, proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra* for all 81 children who participated on both days. On the bottom, results of the cluster based permutation analysis for the same group. The left picture and table show the results from day one, the right picture and table those from day two. The clusters in bold turned out significant after the second step of the analysis.





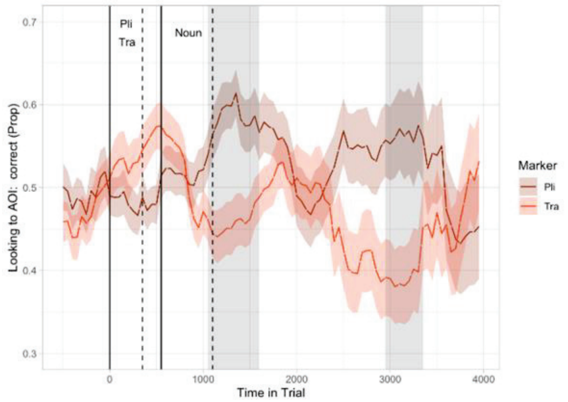
Day 1

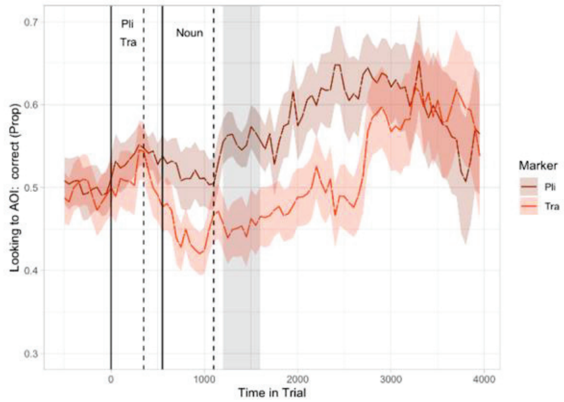
Cluster	Sum statistic	Time range (ms)	*p*-value
1	4.65	3000 – 3100	.499

Day 2

Cluster	Sum statistic	Time range (ms)	*p*-value
1	**28.09**	**950 – 1400**	**.046**

Figure 6.7. Explicitly instructed learners, both days: on the top, proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra* for all explicitly instructed learners who participated on both days. On the bottom, results of the cluster based permutation analysis for the same group. The left picture and table show the results from day one, the right picture and table those from day two. The cluster in bold turned out significant after the second step of the analysis.





Day 1

Cluster	Sum statistic	Time range (ms)	*p*-value
1	-2.02	300 – 350	.731
2	-8.68	400 – 550	.303
3	**33.10**	**1050 – 1600**	**.024**
4	2.15	2550 – 2600	.648
5	**19.88**	**2950 – 3350**	**.092**

Day 2

Cluster	Sum statistic	Time range (ms)	*p*-value
1	2.17	700 – 750	.668
2	**20.71**	**1200 – 1600**	**.078**
3	9.02	2000 – 2200	.310
4	4.70	2250 – 2350	.468
5	5.53	2400 – 2500	.424
6	2.34	2600 – 2650	.608

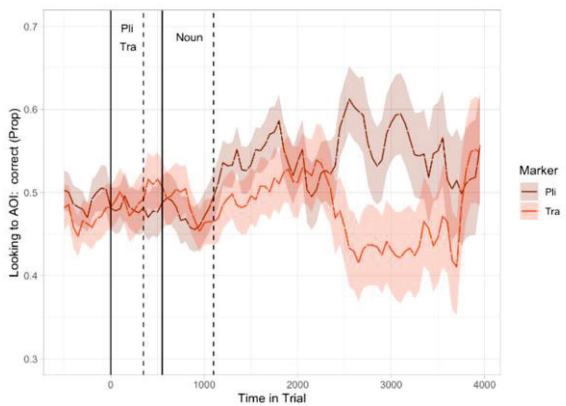
Figure 6.8. Uninstructed learners, both days: on the top, proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra* for all learners who received exposure only and participate on both days. On the bottom, results of the cluster based permutation analysis for the same group. The left picture and table show the results from day one, the right picture and table those from day two. The clusters in bold turned out (marginally) significant after the second step of the analysis.

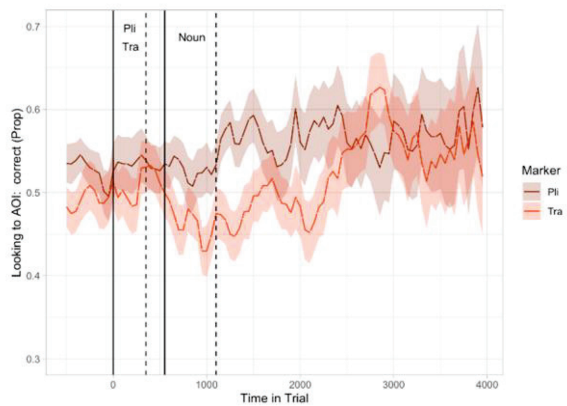
6.2.5 Eye tracking data: Effect of instruction across days

To investigate whether knowledge is consolidated differently for explicitly instructed and uninstructed learners, we split the group based on whether they received explicit instruction or not and compared these groups across the two days. Figures 6.7 and 6.8 show the results for the instructed and uninstructed learners, respectively. For the instructed learners, no effect was visible on day 1. On day 2, however, there was evidence of a preference for looking to *pli* sentences between 950 and 1400 ms. Uninstructed learners looked significantly more often towards the target picture for sentences with *pli* than for sentences with *tra* in two clusters on the first day: one that started when the sentence ended, and a smaller cluster that occurred almost 2 seconds after the end of the sentence. On the second day these uninstructed learners showed predictive eye movements only in one cluster which is comparable to the one we see in the instructed learners on day two.

6.2.6 Eye tracking data: Effect of input across days

A comparison of groups on the basis of whether they received extra input on day two suggests that extra input does not make a difference for eye movement behavior. Figures 6.9 and 6.10 show no difference in eye movement behavior on day 1 and 2. The group of learners that received additional input on day two showed no signs of prediction on day 1 or day 2.





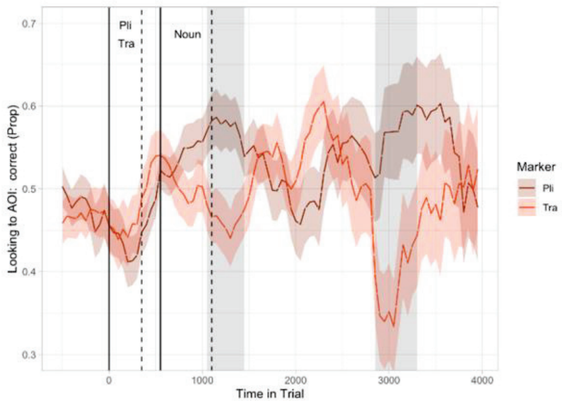
Day 1

Cluster	Sum statistic	Time range (ms)	*p*-value
1	7.80	2550 – 2700	.311
2	4.25	3050 – 3150	.530

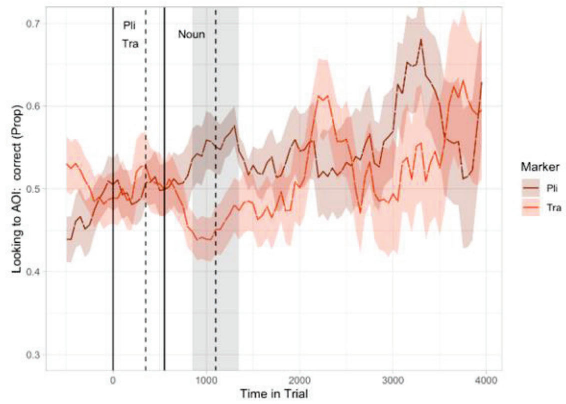
Day 2

Cluster	Sum statistic	Time range (ms)	*p*-value
1	4.21	950 – 1050	.562
2	17.97	1200 – 1550	.126
3	6.96	2050 – 2200	.399

Figure 6.9. On the top, proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra* for learners who participated on both days and received additional input on the second day. On the bottom, results of the cluster based permutation analysis for the same group. The left picture and table show the results from day one, the right picture and table those from day two.



Day 1



Day 2

Cluster	Sum statistic	Time range (ms)	*p*-value
1	**19.07**	**1050 – 1450**	**.078**
2	**22.98**	**2850 – 3300**	**.046**
3	2.40	3650 – 3700	.643

Cluster	Sum statistic	Time range (ms)	*p*-value
1	-4.92	850 – 1350	.477
2	**25.33**	**850 – 1350**	**.046**

Figure 6.10. On the top, proportion of looks towards the target picture over time for sentences with *pli* and sentences with *tra* for learners who participated on both days and did not receive additional input on the second day. On the bottom, results of the cluster based permutation analysis for the same group. The left picture and table show the results from day one, the right picture and table those from day two. The clusters in bold turned out (marginally) significant after the second step of the

6.2.7 Debriefing

In the debriefing, when children were asked how they made a decision on the picture matching task, participants either reported they did not know how they made a decision (*N* = 75) or they claimed that they had heard the sentences before and remembered what they meant (*N* = 6). When asked for the meaning of *pli* and *tra*, most children said they did not know or could not give a meaning (*N* = 67), gave the meaning of a noun they learned during the exposure (*N* = 9), or reported another meaning (*N* = 3), which was not related to the concepts in the experiment. Two children gave answers that were related to number. One child who did not receive explicit instruction, but who did receive extra input on the second day thought that *tra* meant 'two'. Another child who received explicit instruction, and also extra input on the second day answered that *pli* meant 'three' and *tra* meant 'five'.

6.3 Discussion

Studies in second language acquisition consistently show that adults benefit from receiving explicit instruction about the target structures they have to acquire (Goo et al., 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). Yet, whether children are able to make use of explicit instruction is an under researched topic (Lichtman, 2016). To gain more insight into this issue, in the study from Chapter 5, we exposed kindergarteners to a miniature language, from which they had to acquire a grammatical marker indicating number. Results suggested that explicit instruction did not influence accuracy in a picture matching task measuring knowledge of the marker, but that it did lead to earlier predictive eye movements during the task. These results were difficult to interpret, and the current study therefore aimed to investigate possible explanations for these results: we tested children on a second consecutive day to investigate how knowledge develops after a consolidation time including sleep, and after receiving more input.

The findings from the present experiment indicate there was no observable effect of instruction on the accuracy of the grammatical marker on either day of testing. We also did not see an effect of a consolidation period including sleep on children's knowledge of the marker, which has been suggested in the literature (Feld & Diekelman, 2015; Gómez et al., 2006; Spada & Tomita, 2010). We did observe two other interesting patterns. Explicit instruction and consolidation may have affected comprehension of infrequent nouns, as explicitly instructed children gave more target answers for both sentence types on the second day. Second, receiving extra input on the second day did have an effect on accuracy of the grammatical marker. Children who received extra input showed a bigger difference between sentence types on the second day than on the first day in the picture matching task. Yet, we found no observable effect of input on eye movements: there was no evidence that the variable 'input' influenced looking behavior on the second day compared to day one. Interestingly, looking behavior did seem to be influenced by the type of instruction children received. If we take our two studies together, we see that the moment at which explicitly instructed learners show predictive eye movements during the sentence fluctuates. In Chapter 5, children showed predictive looking behavior very early during sentences, whereas in the current study they do not show such predictive looking behavior on the first day of testing, but only on the second day. In short, input seems to have an effect on accuracy scores of the grammatical marker over two days, whereas explicit instruction seems to have some effect on the eye movements of the children, and perhaps on their knowledge of the nouns. The question is how we should explain these results.

One striking observation is that the accuracy scores from the picture matching task and the eye tracking data show different patterns. Many authors have argued these different types of measures tap into different types of knowledge (Andringa et al., 2011; Bowles, 2011; R. Ellis, 2005; Han & Ellis, 1998). Picture matching tasks have features that are typically associated with measures that tap into explicit knowledge (Andringa & Rebuschat, 2015; Godfroid et al., 2015; Loewen, 2009; Sagarra & Seibert Hanson, 2011). Some scholars argue that the strength of a mental representation also determines whether knowledge is available to awareness (Cleeremans, 2008, 2011), and thus whether knowledge is more likely to be explicit or implicit. If we assume this is the case, we could say that good performance on a picture matching task depends on having sufficiently strong linguistic representations, whereas eye movements might be sensitive to knowledge that is still under development. This way, recording eye movements might enable us to measure representations that are still weakly encoded. Following this line of reasoning, the picture matching task and eye tracking results combined might indicate that, during the rule exposure phase, representations of the grammatical rule are constructed and that the transition from less robust to more robust representations is enhanced by additional input: children who received additional input on the second day also performed better during the second picture matching task. We saw an increase in knowledge of the marker for children who received extra input on the second day, as evidenced by the significant three-way interaction between grammatical marker, day of testing and input condition. A consolidation period involving sleep itself does not lead to a learning effect on the second day; only when children receive input again on the second day do they improve their knowledge of the grammatical marker. Such an explanation would be in accordance with the idea that input frequency determines whether grammatical regularities are acquired or not (Ellis, 2002; Bybee, 2010).

As opposed to receiving additional input, explicit instruction in our experiments did not lead to an increase in accuracy of the grammatical marker. In this and in our previous experiment (in Chapter 5) we found no clear evidence that instruction leads to better performance on the picture matching task. In fact, in both studies, we found a small non-significant negative interaction between instruction type and learning the grammatical marker, suggesting that children might be negatively affected by explicit instruction, but this effect is of such small size that we think that even if it were significant, it would be negligible. Yet, we did see a difference in eye movements between instructed and uninstructed children. In all our analyses, we observed that whenever there was a difference in looking behavior between sentences with *pli* and *tra*, this difference was always in the expected direction. For all learners together, this effect was observed for quite a long period (1050 - 1600 ms after the onset of the marker), which is comparable to what we found in the study presented in Chapter 5, where it lasted even longer (700 - 1850 ms after the onset of the marker). Importantly, in our previous study we found that explicitly instructed children showed predictive eye movements early during the sentence on the day they were exposed to the artificial language, while in our current experiment the explicitly instructed group only showed predictive eye movements on the second day, but not on the first day of testing. For uninstructed learners, the predictive eye movements seem stable over the two days: they occurred at similar points in time on both days of testing. The fact that in our two experiments, eye movements of the explicit instructed learners seem to be variable (i.e. predictive eye movements did not occur at similar points in time during the sentence), could indicate that these learners have an unstable low level representation of the grammatical marker. The fluctuating looking behavior in explicitly instructed learners might reflect some sort of doubt about how to process these markers. Such deliberation or variability could be regarded as an early sign of learning (Gilbert & Wilson, 2007; Redish, 2016; Tolman, 1948), and explicit instruction might have fueled this doubt, because children were uncertain about how to use the instruction exactly when learning the markers. Apparently, the explicit instruction about the grammatical element, is not useful enough to result in more stable mental representations. This might suggest that explicit instruction results in unstable mental representation in this age group, which are reflected by their eye movements, but does not help building more robust representations that are needed to show more accuracy of the marker in the in the picture matching task.

Since children first need to determine the word boundaries and deduct what the nouns mean, before they can use statistical dependencies to acquire the grammatical marker (Romberg & Saffran, 2010), acquiring the marker might be relatively difficult for the children, let alone to build sufficiently strong representations that allow for a beneficial use of explicit instruction. If children are indeed making these first steps in acquiring the marker, this might also explain why we consistently observe a relatively small learning effect in studies using the same miniature language. Although the learning effect in this sample was not significant, it was in the same direction and of comparable size to the effect we observed in the study we are trying to replicate from Chapter 5. There, children gave slightly more target answers for sentences with *pli* than for sentences with *tra,* which can be seen as a reflection of acquisition of the grammatical marker. Moreover, in the two other experiments from Chapters 2 and 4 that use the same miniature language, we also observe effects that are in this direction and of a similar size. The results of the current and earlier studies are listed in Table 6.3. Even though the learning effect is not significant in each individual sample, it seems that taken together these studies show that children are, to some extent, able to pick up the grammatical regularity from limited input. Possibly, in these experiments children are building low level representations that can just be measured by an accuracy task, but do not allow for exhaustive use of explicit instruction.

Table 6.3. Results from the generalized logistic effects model for the exposure only group only of our experiment, and studies from the earlier chapters that we were aiming to replicate with this part of the experiment.

Study	***OR***	**95% CI**	***z***	***P***
Chapter 2	1.515	1.071 - 2.257	2.178	.029
Chapter 4	1.344	0.980 - 1.843	1.834	.067
Chapter 5	1.334	0.913 - 1.948	1.490	.136
Current	1.230	0.955 - 1.585	1.602	.109

Furthermore, if children at this stage are also struggling to cope with acquiring the nouns of the language, this could also explain why we found a consolidation effect of the infrequent nouns, which fits well with the idea of how such lexical knowledge is consolidated in this age group (Van den Berghe, 2019). Possibly, the explicit instruction did not call (enough) attention to the marker, but did help children to recognize certain word boundaries: namely that *pli* and *tra* are words in the speech stream, and that the noun that follows is a separate chunk in the input. Instead of helping them to acquire the marker, the explicit instruction may have helped children to become better speech segmenters. This idea is supported by the accuracy data, where we saw that children produced more target answers on the second day than on the first day for both sentences with *pli* and *tra*, but the difference between these conditions did not differ across days. This effect could be driven by the learners who received explicit instruction, as we observed a marginally significant interaction between day of testing and instruction type. This interaction seems to indicate that children who received explicit instruction did not show a significant difference in learning effects between the two days of testing, but did produce more target answers in general on the second day. Giving more target answers for both sentences most likely reflects better knowledge of infrequent nouns, as the nouns are the only information learners can use to give a correct answer on sentences containing *tra*. We might thus argue that explicitly instructed learners have better noun knowledge on the second day than on the first day. Perhaps the explicit instruction did not raise awareness of the grammatical regularity enough to show better performance during a picture matching task, but it did help learners in some way to become better noun learners. Furthermore, children's eye movements might reflect that explicit instruction resulted in unstable representations of the marker.

If young children are able to use explicit instruction to a certain extent, this means they are not necessarily *only* implicit statistical learners. Some authors argue the fundamental difference in learning language between children and adults is that the former only learn implicitly, whereas the latter can learn implicitly as well as explicitly (DeKeyser, 2000, 2003; R. Ellis, 2005, 2009; Paradis, 2004, 2009; Ullman, 2001). Our results could be taken as an indication that this difference is not so much fundamental, but rather gradual. Children are able to process explicit instruction, but just not to the same extent as adults do: in children it might lead to representations that are rather unstable and need additional time or input to become more robust. Explicit instruction is often thought to make learners aware of the target regularities (Hamrick & Rebuschat, 2012). If there is indeed a gradual, and not absolute, difference between adults and children in the extent they can use instruction and become aware of the linguistic structures they need to acquire, this would be in line with theories that assume that awareness exists in different gradations (Allport, 1988; Dehaene et al., 2017; Dennett, 1993; James, 1890). Although children can be made aware of linguistic structures they have to acquire at some level, such awareness is of a lower level than that of adults.

If awareness through explicit instruction indeed plays a role in child language acquisition, its role is most likely subject to individual differences. Studies show that adults can greatly differ in whether they develop awareness of target regularities or not, and whether they make use of such awareness (e.g. Kerz et al., 2017). This study was exploratory and did not have any strong predictions about whether explicit instruction would make a difference for young learners. Therefore, we did not investigate these individual differences, since we were merely interested in seeing whether explicit instruction would have any effect in a random sample of child language learners. Yet, the results of the present study, combined with earlier research (Lichtmann, 2016; Chapter 5 from this book), suggest it is a topic well worth researching further. In line with recent calls for a focus on individual differences in language acquisition studies (Kidd, Donnelly & Christiansen, 2018) further research could investigate how individual differences determine whether children use explicit instruction and develop awareness of the linguistic structures they have to acquire. Such research might shed a better light on how these factors contribute to the language acquisition process. Some other ideas for this will be presented in the next chapter.

Chapter 7

Discussion

In this book, we presented a series of studies that were designed to investigate the contribution of statistical learning to acquiring meaningful grammatical elements and the role of awareness therein. Using a series of miniature language learning experiments, we aimed to answer the following questions:

a) To what extent is statistical learning involved when young children acquire grammatical elements that carry meaning?
b) Can we assess whether young children develop awareness of the knowledge they acquire of such elements?
c) Do young children benefit from being made aware of the grammatical elements they acquire, by receiving explicit instruction about what they have to learn?

Based on the results presented in the previous chapters, a brief answer to each of these questions is given below. In the remainder of this chapter we will try to further contextualize each of these answers, while discussing some of the further methodological and theoretical implications we think these findings might have.

a) Statistical learning supports the acquisition of grammatical elements that carry meaning by young children, but other factors likely contribute to this process too (Chapters 2 & 4-6).
b) With the right methods, it is possible to establish whether children have developed awareness of the knowledge they acquire in this process (Chapters 3 & 4).

c) Being made aware of the target structure through explicit instruction influences the acquisition process, but the precise nature of this influence is unclear (Chapters 5 & 6).

7.1 Statistical learning supports the acquisition of grammatical elements that carry meaning by young children, but other factors likely contribute to this process too

In the experiment we presented first in Chapter 2 and replicated in Chapters 4 to 6, we exposed kindergarteners to a miniature language which contained the grammatical markers *pli* and *tra*. In the language, nouns were always preceded by either one of these markers, which determined whether the noun should be interpreted as plural or not. Whenever children heard *pli* the noun always referred to multiple objects, whereas when *tra* preceded a noun, the noun could refer to a singular object, or to multiple objects. The distributional patterns in the input were the only triggers to acquire this regularity. At test, children heard sentences with *pli* and sentences with *tra,* and had to match these to the right picture. Learning the meaning of the markers should enable them to give correct answers for sentence with *pli*, whereas knowledge of the marker should not be helpful for sentences including *tra*. In each of the experiments we discussed in this book, children showed the following behavior: they consistently gave more correct answers when sentences contained *pli* than when sentences contained *tra*. However, in each of these studies, the effect was small, and it did not reach significance in every individual study. Yet, the results of these experiments all point in the same direction, and taken together, they indicate that children can learn the markers using statistical learning.

The use of the same experimental set up in different samples allows us to go beyond a mere comparison of effect sizes of the separate studies. This enabled us to conduct a small meta-analysis that combines the data from the four different experiments. Using such an approach, we can estimate the size of a possible learning effect for children who are exposed to the language and tested on a single day, more precisely. We used the *metafor* package (Viechtbauer, 2010) to calculate a single average weighted effect size from the experiments in Chapters 2 and 4 to 6. Since there were (minor) methodological differences between our experiments, we assumed that each experiment did not necessarily measure the same underlying true mean. We therefore used a random-effects model, which can be used in such instances (Borenstein, Hedges, Higgins, Rothstein, 2009), and treated the participants from every experimental condition as a separate sample. This means that, for example, learners who received exposure only and those who received exposure and explicit instruction in the experiments from Chapters 5 and 6 were considered to be different samples. Furthermore, from the experiment presented in Chapter 6, we only took results from the picture matching task of the first day of testing. As a result, we ended up with 7 samples from our experiments, and then used the natural logarithm of the *Odds Ratio* of learning for each of these samples in our calculations (Borenstein et al., 2009).

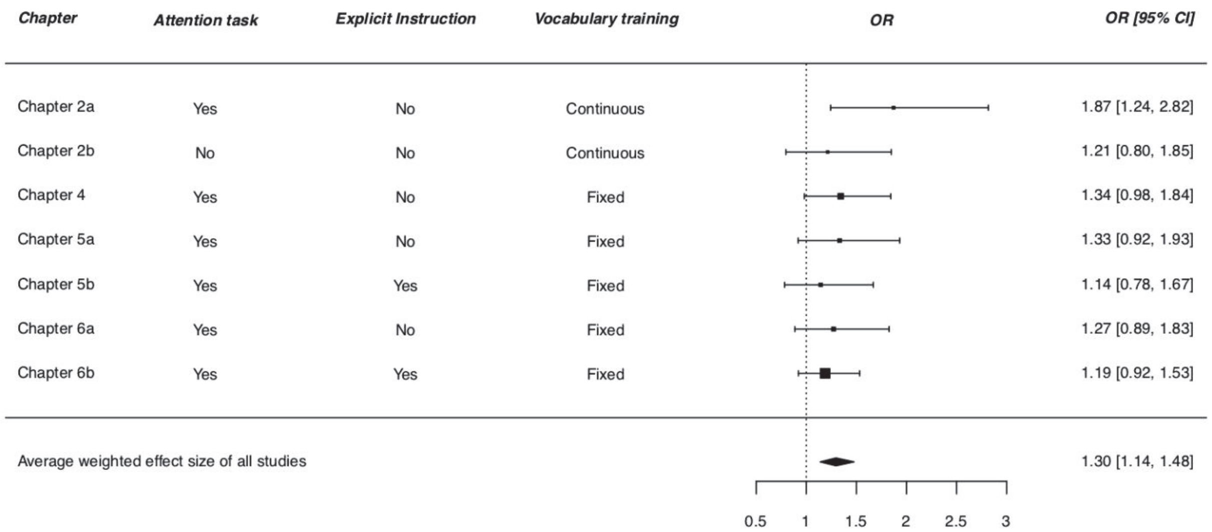


Figure 7.1. A forest plot showing the results from the 7 samples of the experiments presented in this book. The average weighted effect size and its 95% CI are at the bottom.

The results from this meta-analysis can be seen in Figure 7.1, where we observe that, if we take all experiments together, children give more correct answers for sentences including *pli* than *tra,* and that this weighted average effect was significant (average weighted *OR* = 1.30, 95% CI = [1.14…1.48], $p < .001$). Furthermore, Cochran's Q-test showed that heterogeneity between samples was not significant ($Q[6] = 4.133$, $p = .659$), which indicates that despite the minor methodological differences, all samples share a single underlying true effect of learning the marker. On our OSF page (https://osf.io/bp5qe/), we present further meta-regressions in which we investigated whether these methodological differences (attention task, instruction type, vocabulary training) influenced the learning of the number marker. These analyses did not provide any evidence of such effects, which may be due to the low number of samples for certain experimental conditions in our subset. All in all, this small meta-analysis is compatible with the idea that children are able to pick up the meaning of the marker, although the average weighted effect size was still only small.

Based on the outcome of this meta-analysis we can make two important points. First, this analysis underscores the importance of carrying out replications: the combination of the results from these four experiments in which children had to learn the same marker enabled us to get a more precise approximation of the true learning effect than we would have gotten on the basis of, for example, only the results of Chapter 2. The average weighted effect size was a bit smaller than the one we found in our first experiment, and if we had stopped after that experiment, we might have slightly overestimated our estimate of the learning effect. Given that replications are scarce in our field (Marsden et al. 2018), many claims may be based on incorrect effect estimates, and turn out to be misrepresentations after replication (see however Morgan-Short et al., 2018 for one way to tackle this issue, using multi-site replication). This is not only a general problem, but is particularly relevant for the field of statistical learning, where effect sizes are often relatively small (e.g., Kerz et al., 2017), and where there are hints that publication bias might facilitate the overestimation of such small effects (Lammertink et al., 2020; van Witteloostuijn et al., 2017).

Second, this meta-analysis seems to show that if statistical learning is indeed involved in acquiring meaningful grammatical markers, its contribution is only small, at least when the exposure is as brief as in the experiments presented in this book. General claims that statistical learning is the driving force behind language acquisition (Erickson & Thiessen, 2015; Romberg & Saffran, 2010) might need some nuance, because they are generally based on experiments that make use of relatively short exposure. Maybe larger learning effects would be observed if participants received substantially more input. The results presented in Chapter 6 underline this idea: there we saw a slight increase in knowledge of the marker on a second day of testing, but only for learners who received additional input for learning the marker.

In all, we might argue that some sort of association-based learning mechanism is important when children acquire linguistic patterns, but that it is difficult to establish how big its exact contribution is based on experimental set-ups in which children are exposed to the language for relatively brief periods of time. Although it is fascinating to see that children show some learning even after a short exposure, the small effect size shows there is room for other factors to influence statistical learning (Lidz & Gagliardi, 2015; Yang & Montrul, 2017). One of these factors might be awareness of the structures being acquired. Perhaps such awareness interacts with statistical learning mechanisms available to children. However, before investigating whether awareness influences statistical learning, one should know whether children develop awareness of the linguistic elements they need to acquire, and how one can measure such awareness.

7.2 With the right methods, it is possible to establish whether children have developed awareness of the knowledge they acquire in this process

In Chapters 3 and 4, we have tried to make some first steps to enable further research into the role awareness could play in language acquisition. A common assumption is that children do not develop any awareness of the linguistic structures they acquire (Aslin et al., 1998; Chomsky, 1986; Radford, 2004; Saffran et al., 1996; Tomassello, 2003; Ullman, 2016; Wijnen, 2013), and that this assumption might be based on the idea that, when they are aware of something, subjects must be able to verbalize their knowledge (i.e., describe or reflect on it through language). Young children are typically unable to do so, which we also observed in the debriefings in each of our experiments. However, it is conceivable that awareness is dissociated from verbalization (Allport, 1988; Cleeremans, 2008, 2011, 2014; Dehaene et al., 2017; Dennett, 1993; James, 1890). Accordingly, measuring such awareness requires new methods, in which participants are not directly asked to verbalize what they know about the language they are learning (Timmermans & Cleeremans, 2015). In Chapter 3, we introduced such a method: the opt out paradigm. This method allows children to show awareness through a non-verbal response: by opting out. In Chapter 4, we combined this method with our miniature language to see whether children developed awareness of the grammatical marker they had to acquire. The opt out behavior we observed in this experiment suggests that children indeed developed such awareness, even though they were unable to verbalize it.

A first note we should make here, is that we think these findings are not only relevant for scholars working in the field of statistical learning, but also for researchers who approach language acquisition from other theoretical perspectives. Although we have investigated whether children develop awareness of the linguistic structures they are acquiring with the idea in mind that a statistical learning mechanism is involved in this acquisition process, the finding that children develop awareness of their acquired linguistic knowledge is independent of general theoretical assumptions about language acquisition. We approached the issue from a statistical learning perspective, because within this perspective it is most explicitly stated that awareness plays no role in the acquisition process (Conway et al., 2010; Conway & Christiansen, 2006; Kidd, 2011; Perruchet & Pacton, 2006; Turk-Browne et al., 2005). However, scholars working from generative or usage base frameworks have postulated similar ideas (e.g. Chomsky, 1986; Tomassello, 2003). Yet, since our results with regard to the development of awareness are not restricted to a particular theoretical framework, we hope they are of interest to the broader linguistic community as well.

Second, we need to stress that although we think that the results from Chapter 4 show that at the group level children develop awareness of this linguistic pattern, this does not necessarily imply that each member of the sample developed awareness to the same extent. Individual differences likely play a role when children develop awareness of the grammatical structures they acquire, but we merely aimed to see whether such awareness could play a role at all. As such, we did not try to explain why some children would be more likely to develop awareness than others. Readers may have noted that we have provided little information about our sample, apart from their age and gender. We assumed that picking a convenient sample from the population of kindergarteners sufficed to answer the question whether kindergarteners develop the type of awareness we are interested in. There are no a priori reasons to assume that there was a sampling bias that is problematic for the questions that we have tried to answer here, but as a consequence we also did not register any other information about our participants that could provide further insight in other factors that may influence learning, such as their linguistic background or general cognitive capacities (Barac & Bialystok, 2012; Bialystok, 1986). Of course, such information is necessary when we would like to know more about how individual differences in developing awareness relate to differences in, for instance, cognitive or linguistic capacities. The same holds for the samples in Chapters 5 and 6, where we first wanted to investigate whether explicit instruction plays a role in the acquisition of grammatical structures. By resorting to convenience sampling (i.e. those children present in the schools that wanted to participate in these experiments), we ended up with a sample that was representative for the means of our studies (Kruskal & Mosteller, 1979; Kukull & Ganguli, 2012). Future research could further investigate to what extent factors like linguistic background or general cognitive capacities interact with awareness and instruction. The experiments presented in this book were a first step to see whether research on this topic is feasible and worthwhile.

Another factor that could influence whether children develop awareness of the grammatical knowledge they acquire is the nature of the target structure. In this series of experiments, we used a grammatical marker that encoded number. The most important reason for choosing this structure in our experiments was that children seemed to be able to learn i. Note that, unlike other grammatical features (e.g., tense), number has a concrete meaning that can be straightforwardly mapped onto referents in the outside world. Previous studies have also shown that children at this age already possess (some) general numerical knowledge (Meyer, 2019). The concrete and familiar meaning of this number marker might have made it easier for children to develop awareness of it. Natural languages, however, contain grammatical elements with more opaque and less concrete meanings as well (Hengeveld & Mackenzie, 2008). It is not clear whether children would develop awareness of grammatical elements of which the meaning is less concrete, and of which they have less knowledge. Future work is needed to investigate whether children develop awareness of other linguistic structures.

Although it is not clear which cognitive characteristics and linguistic criteria determine whether young children develop awareness of the grammatical structures they acquire, findings like those presented in Chapter 4 raise the possibility of asking new questions. Having established that young children can develop awareness of the grammatical elements they acquire if the right methods are chosen, we can now wonder what possible role such awareness might play within the acquisition process.

7.3 Being made aware of the target structure through explicit instruction influences the acquisition process, but the precise nature of this influence is unclear

Establishing that some children develop awareness when acquiring a grammatical marker opens up the possibility of discussing whether such awareness is helpful in learning or not. In the adult literature this topic has been heavily discussed (DeKeyser, 2003; Ellis, 2003; Krashen, 1981; Schmidt, 1990), but a similar debate is virtually absent in the child language acquisition literature. Meta-analyses on SLA show that adults benefit from being made aware of the target structure through explicit instruction in several linguistic domains (Goo et al., 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). For preschool children, it is often assumed that such explicit instruction does not contribute to the acquisition of grammar (Bialystok, 1994; DeKeyser & Larson-Hall, 2005; Paradis, 2004, 2009). In Chapters 5 and 6, we exposed kindergarteners to the same miniature language, and investigated whether learning the marker is influenced by receiving explicit instruction about it, and whether such learning is a gradual process that develops over time and is strengthened by receiving more input. Children took part in a picture matching task during which eye movements were recorded to test knowledge of the marker. Results from both studies did not provide any evidence that explicit instruction leads to higher accuracy rates, but suggested that it affects children's eye movements. Receiving more input seemingly had the opposite effect: we found that the amount of input that children received positively affected accuracy scores, but did not influence children's eye movements during the task. We argued this might mean that for young learners, explicit instruction leads to rather unstable representations visible in eye movement behavior only, whereas more input leads to more robust knowledge of a grammatical marker in this age group. Explicit instruction seemingly has some effect on young learners, but it seems slightly different from the effect it has on adults. Although it is unclear what the effect precisely might be, these experiments could be considered as a first step into investigating the role explicit instruction and awareness play in child language acquisition. These results could be taken as an indication of the need to pursue this topic.

We should mention here that in the adult language acquisition literature, the focus unsurprisingly lies on second language acquisition. In our experiments, children were also learning a new language, which might raise the question whether these findings are similarly limited to child second language learning, or generalize to first language acquisition. Although learning a second language differs in many respects from learning a first language, several scholars argue there is no fundamental difference in the cognitive capacities that are available for learning these languages at any age (Lichtman 2016; Pakulak & Neville, 2011; Zwart et al., 2017). We similarly think that our findings might also be relevant for first language acquisition: if awareness through explicit instruction plays a role when young children acquire a second language, it might also be involved to a certain extent when they acquire a first language. If the role of awareness and explicit instruction in child language acquisition is indeed investigated further, such research does not necessarily need to be focused so strictly on second language learning. This does not mean the role of awareness and instruction are similar for early first and second language acquisition. Likely, multilingualism and the languages children are learning co-determine the role of awareness within this learning process. We merely argue that explicit instruction and awareness probably play some role in the language acquisition process, whether that is first language acquisition, second language acquisition or multilingual development.

If explicit instruction and awareness indeed play a role in child language acquisition, they most likely interact with many other factors. For adults, we know from meta-analyses about instruction and awareness that factors like length of input, the moment of testing, and the precise linguistic structure all determine the outcome of the learning process (Spada & Tomita, 2010). It is currently impossible to carry out similar meta-analyses with child samples given the gap in the literature. We made a first attempt to fill this void, and investigated some of these factors in the two eye tracking studies reported in Chapter 5 and 6, but these studies only provide us with a very limited view. Most likely explicit instruction, length of input, consolidation, and many other factors determine the language acquisition process for children as well. In line with the idea that language acquisition is a complex dynamic system (Larsen-Freeman, 2012; Yu & Lowie, 2020), we can imagine that the factors we investigated in this book all have a role to play when children acquire languages. In this process, our capacity for statistical learning plays some role, as does awareness of these patterns, but these interact with many other factors and are also subject to individual variation. Future research should investigate how other factors, like linguistic background or general cognitive capacities, interact with explicit instruction in these young learners. Such research would allow us to get a better idea of exactly what roles statistical learning, awareness and explicit instruction play in this age group.

7.4 Concluding remarks

In this book, we aimed to show that statistical learning, awareness, and explicit instruction all come into play when children acquire meaningful grammatical elements. Although the combination of these factors is an under researched topic, we think that the results presented in this book show such research might be promising, and we hope others will continue such investigations. Such research would not only advance our theoretical understanding of this topic, but might also be relevant in more applied settings. Research on similar topics in the field of adult second language acquisition has greatly helped to shape language educational practices. For children, awareness of grammatical structures is sometimes taken into account when developing practical tools for language intervention (see Zwitserlood, Wijnen, van Weerdenburg & Verhoeven, 2015 for an example), but because there is little theoretical knowledge (and few empirical studies) of these issues, practical applications are also relatively scarce. As a consequence, it is difficult to formulate clear practical implications. At the very least, we could say the research presented here not only opens up the possibility to further investigate these issues to improve our theoretical understanding of them, but that we should also pursue more applied studies on the topic. Such research could potentially have a similar impact for child language education about grammatical patterns, as it has done for adult second language teaching in this domain. A first step to enable such developments is to establish that awareness has some role to play in the child language acquisition process. We hope this book serves as a helpful first step and that it will inspire and support future developments in both theoretical and practical directions.

References

Aboh, E.O. (2004). *The Morphosyntax of Complement-Head Sequences: Clause Structures and Word Order Patterns in Kwa.* Oxford: Oxford University Press.

Aboh, E.O., & DeGraff, M. (2014). Some notes on bare noun phrases in Haitian Creole and Gungbe. In T.A. Åfarlí & B. Maehlum (Eds.), *The sociolinguistics of grammar* (pp. 206-236). Amsterdam: John Benjamins Publishing Company.

Agresti, A., & Coull, B.A. (1998). Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, *52*(2), 119-126.

Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory 21*, 435-483.

Alcón Soler, E. (2005). Does Instruction Work for Learning Pragmatics in the EFL Context? *System 33*(3), 417-435.

Allport, A. (1988). What concept of consciousness? In A.J. Marcel & E. Bisiach (Eds.), *Consciousness in Contemporary Science* (pp. 159-182). Oxford: Clarendon Press.

Andringa, S., & Curcic, M. (2015). How Explicit Knowledge Affects Online L2 processing. Evidence from Differential Object Marking Acquisition. *Studies in Language Acquisition 37*, 237-268.

Andringa, S., & Dabrowska, E. (2019). Individual Differences in First and Second Language Ultimate Attainment and Their causes. *Language Learning 69*(1), 5-12.

Andringa, S., de Glopper, C.M., & Hacquebord, H. (2011). Effect of Explicit and Implicit Instruction on Free Written Reponses Task Performance. *Language Learning 61*(3), 868-903.

Andringa, S., & Rebuschat, P. (2015). New Directions in the Study of Implicit and Explicit Learning. An Introduction. *Studies in Language Acquisition 37*, 185-196.

Appah, C.K.I. (2003). *Nominal Derivation in Akan. A Descriptive Analysis* (M.A. Thesis). Norway: Norwegian University of Science and Technology.

Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B*, *372*, 20160058.

Arciuli, J., & Simpson, I.C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286-304.

Aslin, R.N., & Newport, E.L. (2014). Distributional Language learning: Mechanisms and Models of Category Formation. *Language Learning*, *64*(2), 86-105.

Aslin, R.N., Saffran, J.R., & Newport, E.L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological science*, *9*(4), 321-324.

Atagi, N., & Sandhofer, C.M. (2020). Early language environments predict aspects of explicit language awareness development. *Language Learning*, *70*, 464–505.

Axelsson, E. L., Williams, S.E., & Horst, J.S. (2016). The effect of sleep on children's word retention and generalization. *Frontiers in Psychology*, *7*: 1192.

Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics 46*(20), 273-330.

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, *100*: 603-617.

Baguley, T. (2012). *Serious Stats*. Basingstoke: Palgrave Macmillan.

Baker, C.I., Olson, C.R., & Behrmann, M. (2004). Role of Attention and Perceptual Grouping in Visual Statistical Learning. *Psychological Science, 15*(7), 460-466.

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: role of language, cultural background and education. *Child Development*, *83*(2), 413-422.

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Batstone, R. (2002). Making Sense of New Language: A Discourse Perspective. *Language Awareness, 11*, 14-29.

Batterink, L.J., Reber, P.J., Neville, H.J., & Paller, K.A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62-78.

van den Berghe, R. (2019). *Social robots as second-language tutors for young children: Challenges and opportunities* (Doctoral dissertation). Netherlands: Universiteit Utrecht.

Berko, J. (1958). The Child's Learning of English Morphology. *WORD*, *14*(2-3), 150-177.

Bertels, J., Boursain, E., Destrebecqz, A., & Gaillard, V. (2015). Visual statistical learning in children and young adults: how implicit? *Frontiers in Psychology*, *5*, 1-11.

Berwick, R.C., & Chomsky, N. (2015). *Why Only Us: Language and Evolution.* Cambridge, MA: MIT Press.

Bialystok, E. (1986). Factors in the Growth of Linguistic Awareness. *Child Development*, *57*(2), 498-510.

Bialystok, E. (1994). Representation and ways of knowing: three issues in second language acquisition. In N.C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 549-569). San Diego, CA: Academic Press.

Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *International Review of Applied Linguistics*, 43, 319-328.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language LearningI*, 56, 9-49.

Birdsong, D., & Vanhove, J. (2016). Age of second language acquisition: Critical periods and social concerns. In S. Nicoladis & E. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 163-181). Washington, DC: American Psychological Association.

Bley-Vroman, R. (1990). The Logical Problem of Foreign Language Learning. *Linguistic Analysis*, *20*(1-2), 3-49.

Blom, E., & Unsworth, S. (2010). *Experimental Methods in Language Acquisition Research.* Amsterdam: John Benjamins.

Booij, G, & van Santen, A. (2017). *Morfologie. De woordstructuur van het Nederlands. 3e geheel herziene druk.* Amsterdam: Amsterdam University Press.

Borenstein, M.H., Higgins, L.V., & Rothstein, J.P.T. (2009). *Introduction to meta-analysis.* Hoboken, United States: John Wiley & Sons.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech*, *46*(2-3), 217-243.

Bowles, M.A. (2011). Measuring implicit and explicit knowledge. *Studies in Second Language Acquisition*, *33*: 247-271.

Brooks, P.J., & Kempe, V. (2019). More Is More in Language Learning: Reconsidering the Less-Is-More Hypothesis. *Language Learning*, *69*(1), 13-41.

Bybee, J., & McClelland, J.L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, *22*(2-4), 381-410.

Bybee, J. (2010). *Language, usage and cognition.* Cambridge: Cambridge University Press.

Casenhiser, D., & Goldberg, A.E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*(6), 500-508.

Chan, A., Yang, W., Chang, F., & Kidd, E. (2018). Four-year-old Cantonese-speaking children's online processing of relative clauses: a permutation analysis. *Journal of Child Language*, *45*(1), 174-203.

Chen, C.H., Gershkoff-Stowe, L, Wu, C.Y., Cheung, H., & Yu, C. (2017). Tracking Multiple Statistics: Simultaneous Learning of Object names and Categories in English and Mandarin Speakers. *Cognitive Science, 41*(6), 1485-1509.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, N. (1986). *Knowledge of Language. Its Nature, Origin and Use.* New York: NY: Praeger.

Christiansen, M. (2019). Implicit Statistical Learning: A Tale of Two Literatures. *Topics in Cognitive Science*, *11*, 468–481.

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language*, *65*, 128-144.

Cleeremans, A. (2008). Consciousness: the radical plasticity thesis. In R. Banerjee & B.K. Chakrabarti (Eds.), *Progress in Brain Research*, *Volume 168* (pp. 19-33). Amsterdam: Elsevier.

Cleeremans, A. (2011). The radical plasticity thesis: how the brain learns to be conscious. *Frontiers in Psychology, 2*(86), 1-12.

Cleeremans, A. (2014). Connection Conscious and Unconscious Processing. *Cognitive Science, 38*, 1286-1315.

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*, 356-371.

Conway, C.M., & Christiansen, M.H. (2006). Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations. *Psychological Science*, *17*(10), 905-912.

Coyne, M.D., McCoach, D.B., Loftus, S., Zipoli Jr., R., & Kapp, S. (2009). Direct Vocabulary Instruction in Kindergarten: Teaching for Breadth, Versus Depth. *Elementary School Journal*, *110*, 1-48.

Culbertson, J., & Newport., E.L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, *139*, 71-82.

Culbertson, J., & Schuler, K. (2019). Artificial Language Learning in Children. *Annual Review of Linguistics*, *5*, 353-373.

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, *25*, 7-29.

Curcic, M., Andringa, S., & Kuiken, F. (2019). The Role of Awareness and Cognitive Aptitudes in L2 Predictive Language Processing. *Language learning*, *69*(1), 42-71.

Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology*, *2*, 124.

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, 358, 486-492.

DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17*, 379– 410.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499-533.

DeKeyser, R. (2003). Implicit and Explicit learning. In C.J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 312-348). Oxford: Blackwell.

DeKeyser, R. (2009). Cognitive-psychological processes in second language learning. In M.H. Long & C.J. Doughty (Eds.), *Handbook of second language teaching* (pp. 119-138). Oxford: Wiley-Blackwell.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J.F. Kroll & A.M.D. de Groot (Eds), *Handbook of bilingualism psycholinguistic approaches* (pp. 88-108). Oxford: Oxford University Press.

Dennett, D. (1993). *Concsiousness explained.* Penguin UK.

Descartes, R. (1637/1968). *Discourse on Method and the Meditations.* Penguin UK.

Dewar, M., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief Wakeful Resting Boosts New Memories Over the Long Term. *Psychological Science*, *23*(9), 955-960.

Dewar, M., Alber, J., Cowan, N., & Della Sala, S. (2014). Boosting Long-Term Memory via Wakeful Rest: Intentional Rehearsal Is Not Necessary, Consolidation Is Sufficient. *PLoS One*, *9*(10), e109542.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*, 114-126.

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, *13*(5), 309-321.

Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies, 11*, 25-45.

Dienes, Z. (2008). Subjective measures of unconscious knowledge. *Progress in Brain Research, 168*, 49-64.

Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1322-1338.

Dienes, Z., & Berry, D.C. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin and Review, 4*, 3-23.

Dienes, Z., Broadbent, D.E., & Berry, D.C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 875-882.

Dienes, Z., & Fahey, R. (1995). The role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 848-862.

Dink, J.W., & Ferguson, B. (2015). *eyetrackingR; An R Library for Eye-tracking Data Analysis*. Retrieved from http://www.eyetrackingr.com.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.

Doughty, C.J. (2019). Cognitive Language Aptitude. *Language Learning*, *69*(1), 101-126.

Dudai, Y. (2002). Molecular bases of long-term memories: a question of persistence. *Current Opinion in Neurobioly*, *12*, 211–216.

Dudai, Y. (2004). The Neurobiology of Consolidations, Or, How Stable is the Engram? *Annual Review of Psychology*, *55*, 51-86.

Dumay, N., & Gaskell, M.G. (2007). Sleep-Associated Changes in the Mental Representations of Spoken Words. *Psychological Science*, *18*(1), 35-39.

Ehri, L.C., Nunes, S.R., Willows, D.M, Schuster, B.V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, *36*(3), 250-287.

Eigsti, I-M., Zayas, V., Mischel, W., Shoda, Y., Ayduk, O., Dadlani, M.B., Davidson, M.C., Lawrence Aber, J., & Casey, B.J. (2006). Predicting Cognitive Control from Preschool to Late Adolescence and Young Adulthood. *Psychological Science*, *17*(6), 478-484.

Ellis, N.C. (2002). Frequency effects in language processing: A review with implication for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143-188.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C.J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 63-103). Oxford: Blackwell.

Ellis, N.C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*, 305-352.

Ellis, N.C. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. Loew (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA & bilingualism* (pp. 35-47). Washington, DC: Georgetown University Press.

Ellis, N. C. (2015). Implicit and explicit learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3-23). Amsterdam: John Benjamins.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: a psychometric study. *Studies in Second Language Acquisition*, *27*, 141-172.

Ellis, R. (2009). Implicit and explicit learning, knowledge, and instruction. In. R. Ellis, S. Loewen, C. Elder, R. Philp & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 3-25). Buffalo: Multilingual Matters.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339– 368.

Endress, A.D., & Bonatti, L.L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*, 247-299.

Erickson, L.C., & Thiessen, E.D. (2015). Statistical learning of language: Theory, validity and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66-108.

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), 1044-1092.

Feld, G.B., & Diekelmann, S. (2015). Sleep smart - optimizing sleep for declarative learning and memory. *Frontiers in Psychology*, *6*, 622.

Fischer, S., Wilhelm, I., & Born, J. (2007). Developmental differences in sleep's role for implicit off-line learning: comparing children with adults. *Journal of Cognitive Neuroscience*, *19*(2), 214-227.

Foote, A.L., & Crystal, J.D. (2007). Metacognition in the Rat. *Current Biology*, 17, 551-55.

Frank, M.C., Goodmann, N.D. & Tenenbaum, J.B. (2009). Using Speakers' Referential Intentions to Model Early Cross-Situational Word Learning. *Psychological Science, 20*(5), 578-585.

Frank, M.C., Tenenbaum, J.B., & Fernald, A. (2013). Social and Discourse Contributions to the Determination of Reference in Cross-Situational Word Learning. *Language Learning and Development, 9*(1), 1-24.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25-42.

Frost, R.L.A., & Monaghan, P. (2016). Simultaneous segmentation and generalization of non-adjacent dependencies from continuous speech. *Cognition, 147*, 70-74.

Frost, R.L.A., & Monaghan, P. (2017). Sleep-Driven Computations in Speech Processing. *PLoS ONE*, *12*(1), e0169538.

Fujita, K. (2009). Metamemory in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition, 12*, 575-585.

Fukushima, J., Hatta, T., & Fukushima, K. (2000). Development of voluntary control of saccadic eye movements: I. Age-related changes in normal children. *Brain and Development*, *22*(3), 173-180.

Fukuya, Y.J., & Martínez-Flor, A. (2009). The Interactive Effects of Pragmatic-Eliciting Tasks and Pragmatic Instruction. *Foreign Language Annals*, *41*(3), 478-500.

Furnes, B., & Samuelsson, S. (2011) Phonological awareness and rapid automatized naming predicting early development in reading and spelling: Results from a cross-linguistic longitudinal study. *Learning and Individual Differences*, *21*, 85-95.

Gabay, Y., Thiessen, E.D., & Holt, L.L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research, 58*(3), 934-945.

Gass, S., Svetics, I., & Lemelin, S. (2003). Differential Effects of Attention. *Language Learning*, *53*(3), 497-546.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press.

Gilbert, D.T., & Wilson, T.D. (2007). Prospection: Experiencing the Future. *Science*, *317*, 1351-354.

Gillon, G.T. (2018). Phonological Awareness: From Research to Practice. New York: The Guilford Press.

Godfroid, A. (2015). The Effects of Implicit Instruction on Implicit and Explicit Knowledge Development. *Studies in Second Language Acquisition*, *38*, 177-215.

Godfroid, A., Loewen, S., Jung, S., & Park, J-H., Gass, S., & Ellis, R. (2015). Timed and Untimed Grammaticality Judgments Measure Distinct Types of Knowledge: Evidence from Eye-Movement Patterns. *Studies in Second Language Acquisition*, *37*(2), 269-297.

Goldberg, A.E., Casenhiser, D., & Sethuraman, N., (2004). Learning argument structure generalizations. *Cognitive Linguistics*, *15*(3), 289-316.

Goldberg, A.E., Casenhiser, D., & White, T.R. (2007). Constructions as categories of language. *New Ideas in Psychology*, *25*, 70-86.

Gómez, R.L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431-436.

Gómez, R.L., Bootzin, R.B., & Nadel, L. (2006). Naps Promote Abstraction in Language-Learning Infants. *Psychological Science*, *17*(8), 670-674.

Gómez, R.L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109-135.

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443-482). Amsterdam: John Benjamins.

de Graaff, R. (1997). The *Experanto* Experiment. Effects of Explicit Instruction on Second Language Acquisition. *SSLA*, *19*, 249-276.

Granena, G., & Yilmaz, Y. (2019). Corrective Feedback and the Role of Implicit Sequence-Learning Ability in L2 Online Performance. *Language Learning*, *69*(1). 127-156.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31-38.

Hampton, R.R. (2001). Rhesus monkeys know when they remember. *PNAS*, *98*(9), 5359-5362.

Hampton, R.R. (2009). Focusing the uncertainty about nonhuman metacognition. *Comparative Cognition & Behavior Reviews*, *4,* 56-57.

Hamrick, P., & Rebuschat, P. (2012). How implicit is statistical learning? In P. Rebuschat & J.N. Williams (Eds.), *Statistical learning and language acquisition* (pp.365-382). Berlin: Mouton de Gruyter.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, 2, 1-23.

Han, Z., & Finneran, R. (2014). Re-engaging the interface debate: strong, weak, none, or all? *International Journal of Applied Linguistics*, *24*, 370-389.

Hartshorne, J.K., Tenenbaum, J.B., & Pinker, S. (2018). A critical period for language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263-277.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Netherlands: Springer.

Hengeveld, K., & Mackenzie, J.L. (2008). *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford: Oxford University Press.

Hochman, J-R., Endress, A.D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, *115*, 444-457.

Hoff, E. (2012). *Research Methods in Child Language: A Practical Guide*. Oxford: Wiley-Blackwell

Housen, A.M. & Pierrard, M. (2006). Investigating instructed second language acquisition. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 235-269). Amsterdam: Mouton de Gruyter.

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*(175), 1-10.

Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit language learning. *Second Language Acquisition*, *27*, 129-140.

Hulstijn, J. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning. The virtues and pitfalls of a two-system view. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25-46). Amsterdam: John Benjamins.

Jaeger, T. F. Post to HLP/Jaeger lab blog. (2009, May 14). Retrieved from https://hlplab.wordpress.com/2009/05/14/random-effect-structure/

James, W. (1890). *The Principles of Psychology*. New York: Henry Holt and Company.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*. Netherlands: Springer.

Jiménez, L., Mendez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 948.

Johnson, C., & Goswami, U. (2010). Phonological Awareness, Vocabulary, and Reading in Deaf Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *53*, 237-261.

Johnson, J.S., & Newport, E.L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60-99.

Kachergis, G., Yu, C., & Shiffrin, R.M. (2014). Cross-situational word learning is both implicit and strategic. *Frontiers in Psychology, 5*, 588.

Kaiser, E. (2013). Experimental paradigms in psycholinguistics. In R.J. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 135-169). Cambridge: Cambridge University Press.

Ke, S., Miller, R.T., Zhang, D., & Koda, K. (2020). Crosslinguistic sharing of morphological awareness in biliteracy development: A systematic review and meta-analysis of correlation coefficients. *Language Learning*. https://doi.org/10.1111/lang.12429

Kerz, E., Wiechmann, D., & Riedel, F.B. (2017). Implicit Learning in the Crowd. Investigating the Role of Awareness in the Acquisition of L2 Knowledge. *Studies in Second Language Acquisition*, *39*, 711-734.

Kidd, E. (2011). Implicit Statistical Learning Is Directly Associated with the Acquisition of Syntax. *Developmental Psychology*, *48*(1), 171–184.

Kidd, E., Donnell, S., & Christiansen, M.H. (2018). *Trends in Cognitive Neuroscience*, *22*(2), 154-169.

Kissling, E.M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal*, *97*(3), 720-744.

Koike, D.A., & Pearson, L. (2005). The Effect of Instruction and Feedback in the Development of Pragmatic Competence. *System*, *33*(3), 481-501.

Krashen, S.D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S.D., Scarcella, R.C., & Long, M.A. (1982). *Child-adult differences in second language acquisition*. Rowley, MA: Newbury House.

Kruskal, W., & Mosteller, F. (1979) Representative sampling: 3: current statistical literature. *International Statistical Review*, *47*, 245-265.

Kukull, W., & Ganguli, M. (2012). Generalizability. The trees, the forest, and the low-hanging fruit. *Neurology*, *78*, 1886-1891.

Lamme, V.A.F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*, 204-220.

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*, 3474-3486.

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.

Lany, J. (2014). Judging words by their covers and the company they keep: Probabilistic cues support word learning. *Child Development, 85*, 1727-1739.

Lany, J., & Saffran, J.R. (2013). Statistical learning mechanisms in infancy. In J.L.R. Rubenstein & P. Rakic (Eds.), *Comprehensive Developmental Neuroscience: Neurel Circuit Development and Function in the Brain, volume 3* (pp.231-248). Amsterdam: Elsevier.

Larsen-Freeman, D. (2012). Complex, dynamic systems: A new transdisciplinary theme for applied linguistics? *Language Teaching*, *45*(2), 202-214.

Léger, P-M., Sénécal, S., Karpova, E., Brieugne, D, Di Fabio, M-L., & Georges, V. (2018). Setup guidelines for eye tracking in child and teenager research in the context of learning by interaction with a tablet. *Neuroeducation*, *5*(1), 33-40.

Lichtman, K. (2016). Age and learning environment: Are children implicit second language learners? *Journal of Child Language*, *43*, 707-730.

Lidz, J., & Gagliardi, A. (2015). How Nature Meets Nurture: Universal Grammar and Statistical Learning. *Annual Review of Linguistics 2015, 1*(1), 333-353.

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, H. Reinders, R. Erlam, J. Philip (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 94-112). Bristol: Multilingual Matters.

Lowie, W.M., & Verspoor, M.H. (2019). Individual Differences and the Ergodicity Problem. *Language Learning*, *69*(1), 184-206.

Lum, J.A.G., Conti-Ramsden, G., Page, D., & Ullman, M.T., (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*, 1138-1154.

Maquet, P. (2001). The role of sleep in learning and memory. *Science*, *294*, 1048-1052.

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental Methods in Language Acquisition Research* (pp. 139-162). Amsterdam: John Benjamins.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *38*(4), 843-847.

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in Second Language Research: Narrative and Systematic Reviews and Recommendations for the Field. *Language Learning*, *68*, 321-391.

Martínez-Flor, A., & Fukuya., Y.J. (2005). The effects of instruction on learners' production of appropriate and accurate suggestions. *System*, *33*(3), 463-480.

Marulis, L.M., & Neuman, S.B. (2010). The Effects of Vocabulary Intervention on Young Children's Word Learning: A Meta-Analysis. *Review of Educational Research*, *80*, 300-335.

Mednick, S.C, Cai, D.J., Shuman, T., Anagnostaras, S., & Wixted, J. (2011). An opportunistic theory of cellular and systems consolidation. *Trends in Neuroscience*, *34*(10), 504-514.

Merikle, P.M., Smilek, D., & Eastwood, J.D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition*, *79*, 115-134.

Meyer, C. (2019). *Rule and Order* (Doctoral dissertation). Netherlands: University of Amsterdam.

Mintz, T.H., Newport, E.L., & Bever, T.G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*, 393-424.

Mirković, J., & Gaskell, M.G. (2016). Does Sleep Improve Your Grammar? Preferential Consolidation of Arbitrary Components of New Linguistic Knowledge. *PLoS ONE*, *11*(4), e0152489.

Mischel, W., Shoda, Y., & Rodriguez, M.L. (1989). Delay of Gratification in Children. *Science* 244: 933-938

Misyak, J.B., & Christiansen, M.H. (2012). Statistical Learning and Language: An Individual Differences Study. *Language Learning*, *62*(1), 302-331.

Misyak, J.B., Christiansen, M.H., & Tomblin, J.B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*, 1-9.

Monaghan, P., Mattock, K., Davies, R.A.I., & Smith, A.C., (2015). Gavagai Is as Gavagai Does: Learning Nouns and Verbs from Cross-Situational Statistics. *Cognitive Science, 39*, 1099-1112.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M.T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*(2), 933-947.

Muranoi, H. (2000). Focus on form through interaction enhancement: Integration formal instruction into a communicative task in EFL classrooms. *Language Learning*, *50*(4), 617-673.

Newport, E.L. (2020). Children and Adults as Language Learners: Rules, Variation, and Maturational Change. *Topics in Cognitive Science*, *12*(1), 153-169.

Nieuwenhuis, I.L.C., Folia, V., Forkstam, C., Jensen, O., & Petersson, K.M. (2013). Sleep Promotes the Extraction of Grammatical Rules. *PLoS One*, *8*(6), e65046.

Norris, J.M., & Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning*, *50*(3), 417-528.

Packard, M.R. (2009). Anxiety, cognition and habit: A multiple memory systems perspective. *Brain Research*, *1293,* 121-128.

Pakuklak, E., & Neville, H.J. (2011). Maturational Constraints on the Recruitment of Early Processes for Syntactic Processing. *Journal of Cognitive Neuroscience*, *23*(10), 2752-2765.

Pan, B.A. (2012). Assessing Vocabulary Skills. In E. Hoff (Ed.), *Research Methods in Child Language: A Practical Guide* (pp. 100-112). Oxford: Wiley-Blackwell.

Paradis, M. (2004). *A neurolinguistics theory of bilingualism.* Amsterdam; John Benjamins.

Paradis, M. (2009). *Declarative and Procedural Determinants of Second Languages*. Amsterdam: John Benjamins.

Parnes, A.T. (2011). Ang *marks the what? An analysis of noun phrase markers in Cebuano* (M.A. Thesis). New Have, CT: Yale University.

Passer, M. (2016). *The Typology and Diachrony of Nominal Classification* (Doctoral dissertation). Netherlands: University of Amsterdam.

Perruchet, P., & Pacton, S., (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Science*, *10*(5), 233-238.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for Word Segmentation. *Journal of Memory and Language, 39*, 246-263.

Pinker, S. (1994). *The Language Instinct.* New York: Penguin Books.

Pinker, S. (1997). *How the mind works.* New York: Penguin Books.

Pinto, M., & Zuckerman, S. (2019). Coloring Book: A new method for testing language comprehension. *Behavior Research Methods*, *51*, 2609-2628.

Podesva, R.J., & Sharma, D. (2013). *Research Methods in Linguistics*. Cambridge: Cambridge University Press.

Powell, M.J.D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives.* Technical Report, Department of Applied Mathematics and Theoretical Physics, University of Cambridge.

Proust, J. (2013). *The philosophy of metacognition*. Oxford, UK: Oxford University Press.

Psychology Software Tools. (2012). E-Prime 3.0 [Computer software]. Retrieved from http://www.pstnet.com.

Qian, T., Reeder, P.A., Aslin, R.N., Tenenbaum, J.B., & Newport, E.L. (2012). Exploring the Role of Representation in Models of Grammatical Category

Acquisition. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*, 881-886.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Radford, A. (2004). *English Syntax: An Introduction.* Cambridge: Cambridge University Press.

Raviv, L., & Arnon, I. (2018). Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition*, *181*, 160-173.

Reber, A. (1967). Implicit Learning of Artificial Grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855-863.

Rebuschat, P. (2013). Measuring Implicit and Explicit Knowledge in Second Language Research. *Language Learning, 63*(3), 595-626.

Rebuschat. P., & Williams, J.N. (2012). Introduction: Statistical learning and language acquisition. In P. Rebuschat & J.N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 1-12). Berlin: Mouton de Gruyter.

Redish, A.D. (2016). Vicarious trial and error. *Nature Reviews Neuroscience*, *17*(3), 147-159.

Reed, J., & Johnson, P. (1994) Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 585.

Reingold, E.M., & Merikle, P.M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*, *44*, 563-575.

Reingold, E.M., & Merikle, P.M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language*, *5*, 10-28.

Reingold, E.M., & Toth, J.P. (1996). Progress dissociations versus task dissociations: A controversy in progress. In G. Underwood (Ed.) *Implicit cognition* (pp. 159-202). Oxford: Oxford University Press.

Robinson, P. (1997a). Generalizability and automaticity of second language learning under implicit, incidental, enhanced and instructed conditions. *Studies in Second Language Acquisition*, *19*(2), 233– 247.

Robinson, P. (1997b). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*, 45– 99.

Romberg, A.R., & Saffran, J.R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906-914.

Saffran, J.R., Johnson, E.K., & Aslin, R.N. (1996). Word-segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Sagarra, N., & Hanson, A.S. (2011). Eyetracking Methodology: A User's Guide for Linguistic Research. *Studies in Hispanic and Lusophone Linguistics*, *4*(2), 1-13.

Saito, K. (2013). The Acquisitional Value of Recasts in Instructed Second Language Speech learning: Teaching the Perception and Production of English /r/ to Adult Japanese Learners. *Language Learning*, *63*(3), 499-529.

Saygin, Z.M., Norton, E.S., Osher, D.E., Beach, S.D., Cyr, A.B., Ozernov-Palchik, O., Yendiki, A., Fischl, B., Gaab, N., & Gabrieli, J.D.E. (2013). Tracking the Roots of Reading Ability: White Matter Volume and Integrity Correlate with Phonological Awareness in Prereading and Early-Reading Kindergarten Children. *Journal of Neuroscience, 33*, 13251-13258.

Schmidt, C., & Miller, K. (2010). Using comprehension methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental Methods in Language Acquisition Research* (pp. 35-56). Amsterdam: John Benjamins.

Schmidt, R.W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, *11*(2), 129-158.

Schütze, C.T., & Sprouse, J. (2013). Judgment data. In R.J. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 27-51). Cambridge: Cambridge University Press.

Scott, V. (1989). An empirical study of explicit and implicit teaching strategies in French. *Modern Language Journal*, *72*, 14– 22.

Sedivy, J.C. (2010). Using eyetracking in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental Methods in Language Acquisition Research* (pp. 115-138). Amsterdam: John Benjamins.

Sheen, Y. (2007). The effects of corrective feedback, language aptitude, and learner attitudes on the acquisition of English articles. In Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 301-322). New York: Oxford University Press.

Singleton, D., & Munoz, C. (2011). Around and beyond the critical period hypothesis. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning, Volume 2* (pp. 407-425). London: Routledge.

Singleton, D., & Ryan, L. (2004). *Language acquisition the age factor. Second edition*. Bristol: Multilingual Matters Ltd.

Silverman, R. (2007). A Comparison of Three Methods of Vocabulary Instruction during Read-Alouds in Kindergarten. *The Elementary School Journal, 108*, 97-113.

Smith, J.D. (2009). The study of animal metacognition. *Trends in cognitive science*, *13*, 389-396.

Smith, J.D. (2010). Inaugurating the Study of Animal Metacognition. *Journal of Comparative Psychology*, *23*, 401-413.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 444-449.

Smith, L.B., Suanda, S.H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Science, 18*(5), 251-258.

Smith, L.B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558-1568.

Sole, L.M., Shettleworth, S.J., & Bennett, P.J. (2003). Uncertainty in pigeons. *Psychonomic Bulletin & Review*, *10*, 738-745.

Spada, N., & Tomita, Y. (2010). Interactions Between Type of Instruction and Type of Language Feature: A Meta-Analysis. *Language Learning*, *60*(2), 263-308.

Spit, S, Andringa, S., Rispens, J., & Aboh, E.O. (2019). The opt out paradigm. First steps towards a new experimental method that measures meta-linguistic awareness. *Dutch Journal of Applied Linguistics*, *8*, 206-227.

Spit, S, Andringa, S., Rispens, J., & Aboh, E.O. (2021). Do kindergarteners develop awareness of the statistical regularities they acquire? *Language Learning*, *71*(2), 573-611.

Spit, S, Andringa, S., Rispens, J., & Aboh, E.O. (2021). The effect of explicit instruction on implicit and explicit linguistic knowledge in kindergarteners. *Language Learning and Development*, https://doi.org/10.1080/15475441.2021.1941968

Spit, S, Andringa, S., Rispens, J., & Aboh, E.O. (under review). Kindergarteners use cross-situational statistics to infer the meaning of grammatical elements. *Journal of Psycholinguistic Research*.

Spit, S, Andringa, S., Rispens, J., & Aboh, E.O. (under review). The effects of instruction, input, and sleep during kindergartener's acquisition of grammatical structures: an eye tracking study. *Language Development Research*.

Spit, S, Andringa, S., Rispens, J., Aboh, E.O, & Vet, D.J. (2019, November 11). Meta-linguistic awareness in early (second) language acquisition. Retrieved from osf.io/bp5qe. https://doi.org/10.17605/OSF.IO/BP5QE

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*, 1272-1278.

Stickgold, R., & Walker, M.P. (2013). Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience*, *16*, 139-145.

Suda-King, C. (2008). Do orangutans (*Pongo pygmaeus*) know when they do not remember? *Animal Cognition*, *11*, 21-42.

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitude. *Language learning*, *67*, 747-790.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*, 86-132.

Tal, S., & Arnon, I. (2020). Do children preferentially mark unpredictable material? The case of optional plural marking. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society,* 3226-3232.

Thiessen, E.D., Kronstein, A.T., & Hufnagle, D.G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin, 139*(4), 792-814.

Timmermans, B., & Cleeremans, A. (2015). How can we measure awareness? An overview of current methods. In M. Overgaard (Ed.), *Behavioural Methods in Consciousness Research* (pp. 21-46). Oxford: Oxford University Press.

Tolman. (1948). Cognitive Maps in Rats and Men, *Psychological Review*, *55*, 189-208.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1-2), 61-82.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press

Toro, J.M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*(2), 25-34.

Turk-Browne, N.B., Jungé, J., & Scholl, B.J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology*, *134*(4), 552-564.

Ullman, M. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Neuroscience*, *2*, 717-726.

Ullman, M. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, *92*, 231-270.

Ullman, M. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge and use. In G. Hickok & S. A. Small (Eds.), *The Neurobiology of Language* (pp. 953-968). Amsterdam: Elsevier.

Ullman, M., & Pierpont, E. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex*, *41*(3), 399-433.

Vaahtoranta, E., Suggate, S., Jachmann, C., Lenhart, J., & Lenhard, W. (2018). Can explaining less be more? Enhancing vocabulary through explicit versus elaborative storytelling. *First language*, *38*, 198-217.

Vichtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software*, *36*(3), 1-48.

Vlach, H.A., & Johnson, S.P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition, 127*, 375-382.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition, 107*, 729-274.

Vouloumanos, A. (2018). Voulez-vous jouer avec moi? Twelve-months-olds understand that foreign languages can communicate. *Cognition*, *174*(4), 87-92.

de Waal, F. (2016). *Are We Smart Enough to Know How Smart Animals Are?* New York, NY: W.W. Norton & Company.

West, G., Vadillo, M.A., Shanks, D.R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Developmental Science*, *21*(2), e12552.

Wijnen, F. (2013). Acquisition of linguistic categories: cross domain convergences. In J. Bolhuis & M. Everaert (Eds.), *Birdsong, Speech, and Language* (pp. 157-177). Cambridge, MA: MIT Press.

Wilhelm, I. Diekelmann, S., & Born, J. (2008). Sleep in children improves memory performance on declarative but not procedural tasks. *Learning & Memory*, *15*(5), 373-377.

Williams, J.N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269-304.

Williams, J.N. (2009). Implicit learning in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319-353). Bingley, UK: Emerald Press.

Williams, J.N., & Evans, J. (1998). What kind of focus and on which forms? In C. Doughty & J.N. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 139-155). Cambridge: Cambridge University Press.

Williams, S.E., & Horst, J.S. (2014). Goodnight book: sleep consolidation improves word learning via storybooks. *Frontiers in psychology*, *5*, 184.

van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2017). Visual artificial grammar learning in dyslexia: A meta-analysis. *Research in developmental disabilities, 70*, 126-137.

Wixted, J. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*, 235-269.

Wonnacott, E., Boyd, J.K., Thomson, J., & Goldberg, A.E. (2012). Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Langauge*, *66*(3), 458-478.

Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language*, *95*, 36-48.

Yang, C., & Montrul, S. (2017). Learning datives: The Tolerance Principle in monolingual and bilingual acquisition. *Second Language Research*, *33*(1), 119-144.

Yu, C., & Smith, LB. (2007). Rapid Word Learning Under Uncertainty via Cross-Situational Statistics. *Psychological Science*, *18*(5), 414-420.

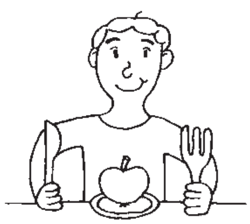
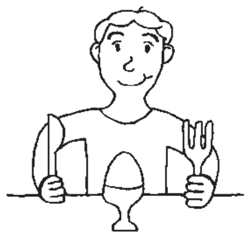
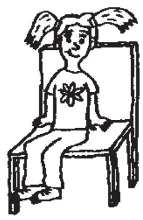
Yu, H., & Lowie, W. (2020). Dynamic Paths of Complexity and Accuracy in Second Language Speech: A Longitudinal Case Study of Chinese Learners. *Applied Linguistics*, *41*(6), 855-877.

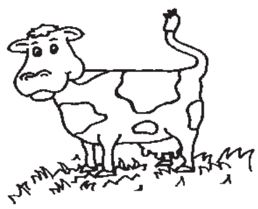
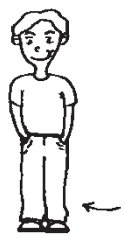
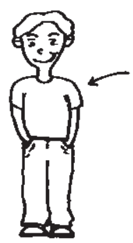
Zwart, F.S., Vissers, C.Th.W.M, Kessels, R.P.C., & Maes, J.H.R. (2019). Procedural learning across the lifespan: A systematic review with implications for atypical development. *Journal of Neuropsychology*, *13*(2), 149-182.

Zwitserlood, R., Wijnen, F., van Weerdenbrug, M., & Verhoeven, L. (2015). 'MetaTaal': enhancing complex syntax in children with specific language impairment – a metalinguistic and multimodal approach. *International Journal of Language & Communication Disorders*, *50*(3), 273-297.

Appendix

All test items and the pictures between which a participant had to choose, when hearing one of the six test items next to them during the experiment presented in Chapter 3.

Test items	Picture pair	
De jongen eet de appel *De jongen eet de pato* *De jongen eet een pato* *De jongen eet het ei* *De jongen eet het domo* *De jongen eet een domo* 'The boy eats the apple/egg'		
Het meisje zit op de stoel *Het meisje zit op de herbi* *Het meisje zit op een herbi* *Het meisje zit op het bed* *Het meisje zit op het glavi* *Het meisje zit op een glavi* 'The girl sits on the chair/bed'		
De vrouw leest de krant *De vrouw leest de fundi* *De vrouw leest een fundi* *De vrouw leest het boek* *De vrouw leest het agri* *De vrouw leest een agri* 'The women reads the paper/book'		

In het gras staat de koe *In het gras staat de lino* *In het gras staat een lino* *In het gras staat het paard* *In het gras staat het orbo* *In het gras staat een orbo* 'The cow/horse is standing in the grass'		
De vogel vliegt naast de boom *De vogel vliegt naast de zumu* *De vogel vliegt naast een zumu* *De vogel vliegt naast het huis* *De vogel vliegt naast het teku* *De vogel vliegt naast een teku* 'The bird is flying next to the tree/house'		
De man draagt de broek *De man draagt de tigru* *De man draagt een tigru* *De man draagt het shirt* *De man draagt het wolgu* *De man draagt een wolgu* 'The man is wearing the shirt/pants'		

Summary

Awareness and instruction when kindergarteners acquire grammar

Acquiring a language encompasses several seemingly distinct tasks: learners need to detect words within a continuous stream of speech, they need to connect meanings to these words and they have to group these words into abstract categories (i.e., nouns and verbs), and determine the grammatical relations between these categories. In one view of how learners accomplish these tasks, acquiring language relies on a general cognitive capacity to detect distributional properties within language, which is often referred to as 'statistical learning'. Statistical learning can be seen as a cognitive association mechanism that operates on different stimuli, and enables learners to create mental representations about patterns in the input, which in turn are used to generate relevant behavioral output. In this book, we present a series of studies that were designed to investigate the contribution of statistical learning to acquiring meaningful grammatical elements in kindergarteners, and the role of awareness therein. These studies are further contextualized in **Chapter 1**.

In **Chapter 2**, we describe the statistical learning experiment that was used in the majority of these studies. In this experiment, participants had to learn a new language: they saw pictures and heard things that matched the pictures they saw. From this input, children had to acquire a plural marker. Afterwards, they were tested by means of a picture matching task to determine whether they had become sensitive to this grammatical structure. At test, participants heard sentences from the language and had to choose which of two pictures matched each sentence. Using this task, we could measure how well they had learned the plural marker. Results from 50 kindergarteners presented in this chapter show

that statistical learning supports the acquisition of grammatical elements by young children, but that other factors likely contribute to this process too.

One of these other factors that we further investigated is awareness of the target structure. Awareness has been studied extensively within the domain of second language acquisition, but very little research has focused on whether young children develop awareness of what they are learning. However, research with adults has also shown that measuring awareness is a difficult enterprise, especially when participants are unlikely to be able to verbalize their awareness. In **Chapter 3**, we present a method with which we tried to overcome this problem: the opt out paradigm. The opt out paradigm assesses awareness without requiring learners to verbalize their experiences. Instead, participants show strategic behavior that reveals their awareness. The results from this chapter suggested that the opt-out method can indeed be used to gain more insight into whether young learners, who are unable to verbalize awareness of acquired linguistic structure, could still possess such awareness. In **Chapter 4,** we combined this method with the earlier described statistical learning experiment, to see whether children develop any awareness of what they learn during such an experiment. The results from this experiment with 70 kindergarteners showed that even though young children cannot verbalize their awareness of acquired linguistic knowledge, they nevertheless may be aware of it.

Following up on these experiments, we wanted know whether children might also benefit from being made aware of the target structure. In two experiments we taught children the same miniature language, and provided half of the children with explicit instruction about the number marker. During the picture matching task that was used to measure learning, we also tracked children's eye movements to investigate whether these would reveal anything about their acquired knowledge. In **Chapter 5**, we present a study in which 113 children were tested using this set up, but did not find any evidence that explicit instruction leads to higher accuracy rates. We did, however, find evidence of changed eye movement behavior because of the explicit instruction that children received. **Chapter 6** presents a replication of this study, but where 103 children were tested twice: once directly after the exposure, and a second time one day later. We did this to get a clearer picture of how explicit instruction affects acquisition of a grammatical marker, as it may take some time before the instruction sinks in. Again, the results did not provide any evidence that explicit instruction leads to higher accuracy rates, but there was evidence of changed eye movement behavior because of such instruction. Importantly, when children received additional input on the second day, their accuracy did increase slightly.

Even though not all results from the experiments presented in this book are conclusive, we aimed to show that the role awareness plays in early language acquisition is a highly relevant, yet under-investigated topic. Statistical learning, awareness, and explicit instruction probably all come into play when children acquire grammatical structures. Although the combination of these factors is an under researched topic, we argue in **Chapter 7** that research into these topics is promising, and we call for more of such investigations.

Samenvatting

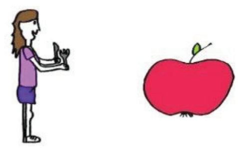
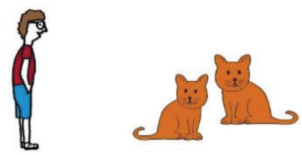
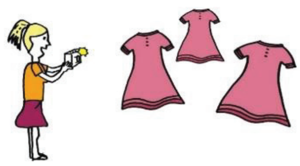
Bewustzijn en instructie bij de taalverwerving van kleuters

Om een taal te leren moeten kinderen van alles kunnen: ze moeten de woorden herkennen in een stroom van spraakklanken en de betekenis van deze woorden leren. Vervolgens moeten ze leren dat bepaalde woorden tot dezelfde grammaticale categorie behoren; sommige woorden zijn bijvoorbeeld lidwoorden ('de', 'het'), terwijl andere woorden zelfstandig naamwoorden zijn ('koe', 'paard'). Daarnaast moeten ze er ook achter komen dat er relaties zijn tussen al die grammaticale categorieën, bijvoorbeeld dat in het Nederlands voor een zelfstandig naamwoord vaak een lidwoord komt ('de koe', 'het paard'). Er zijn verschillende ideeën over hoe taalleerders al die kennis opdoen. In één theorie wordt uitgegaan van een cognitieve capaciteit (dat wil zeggen: vermogen van het brein om allerlei handelingen uit te voeren) die mensen in staat stelt om de statistische patronen in taal te herkennen. Dit vermogen wordt ook wel 'statistisch leren' genoemd. Omdat mensen zo goed zijn in statistisch leren kunnen we allerlei klanken en woorden met elkaar associëren. Met behulp van dat cognitieve vermogen leren we onder andere dat lidwoorden en zelfstandig naamwoorden bij elkaar horen.

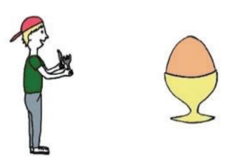
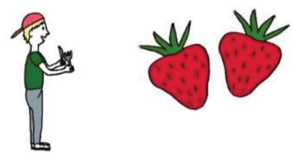
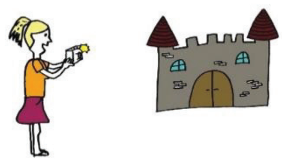
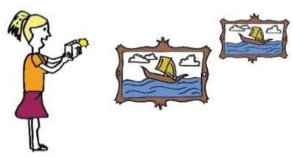
In **Hoofdstuk 1** van dit boek leg ik uit dat er nog veel onduidelijk is over hoe dat statistisch leren precies in zijn werk gaat. Vaak gaan onderzoekers er namelijk van uit dat statisch leren iets is dat impliciet gebeurt: kinderen zijn zich er niet bewust van dat ze aan het leren zijn, en ze weten ook niet wat ze uiteindelijk geleerd hebben. Omdat dit allemaal zo onbewust gebeurt, heeft het ook geen zin om jonge kinderen uitleg te geven. Die uitleg pikken ze waarschijnlijk toch niet op! In dit promotieonderzoek heb ik gekeken of dat klopt. Zijn kinderen zich echt niet bewust van de taalregels die ze leren? En heeft het inderdaad geen zin om ze uitleg te geven? Om een antwoord op die vragen te geven, lieten we - ik heb het vaak over ‘we’, omdat 'ik' dit onderzoek natuurlijk niet in mijn eentje heb uitgevoerd - kleuters als spelletje een kunsttaal leren. Dat is een taal die zelf verzonnen is, en niet door mensen wordt gesproken. Tijdens het experiment zien kinderen allerlei plaatjes en horen zinnetjes in de nieuwe taal die de plaatjes omschrijven. Door zo’n kunsttaal te gebruiken, kun je zelf bedenken welke regels je de kinderen laat leren en maken ze (hopelijk) allemaal hetzelfde leerproces door. Zo achterhaal je dus precies waar de kinderen wel en niet toe in staat zijn. In dit boek zijn verschillende varianten van zo’n experiment gebruikt om erachter te komen of kinderen zich bewust zijn van de taal die ze leren, en of het zin heeft om ze uitleg te geven over de regels in die taal.

In **Hoofdstuk 2** beschrijf ik de basisvariant van het experiment. In het experiment helpen de kleuters vier personages. Die vier personages gaan op vakantie naar een land ver weg en de kleuters moeten met hen een nieuwe taal leren. Om die taal te leren zien de kinderen dus steeds een plaatje en horen iets in de taal dat het plaatje omschrijft. In Tabel S1 op de volgende pagina zie je twee van zulke plaatjes, met de twee (in dit geval geschreven) zinnetjes die daarbij horen. Zoals je op de plaatjes kunt zien, staan naast de personages ook een aantal objecten of dieren. Die objecten of dieren komen in verschillende hoeveelheden voor. De hoeveelheid wordt bepaald door een specifiek woordje in de zinnetjes. In het ene zinnetje staat namelijk het woordje *pli* en in het andere zinnetje het woordje *tra*. De regel is dat wanneer *pli* in het zinnetje staat, er altijd meerdere objecten of dieren te zien zijn. Wanneer *tra* in het zinnetje staat, kan ieder aantal objecten of dieren te zien zijn.

Tabel S1. Zinnetjes en plaatjes die kinderen hoorden tijdens het experiment.

Zin	Plaatje
Maria rigarda tra zambo	
Carlo estima tra pano	
Julia pentura pli anso	

Tabel S2. Voorbeelden van testzinnetjes.

	Marco rigarda pli bovo	
Pli		
	Julia pentura tra nutro	
Tra		

Noot: Voor zinnetjes met *pli* kunnen kinderen leren dat er altijd meerdere objecten op het plaatje moeten staan. Voor zinnetjes met *tra* is het moeilijk een goed antwoord te geven als je niet alle woorden uit de taal kent. In allebei de voorbeelden is het rechter plaatje het goede antwoord.

Om deze meervoudsregel te leren, krijgen de kinderen ongeveer 20 minuten lang plaatjes te zien en bijbehorende zinnetjes te horen. Na afloop van die 20 minuten worden kinderen getest om te kijken of ze deze regel geleerd hebben. Ze horen dan weer een zinnetje in de taal en krijgen twee plaatjes te zien, zoals je in Tabel S2 hierboven kunt zien. Ze moeten vervolgens kiezen welk plaatje bij het zinnetje hoort. Op het ene plaatje staat altijd één object, op het andere plaatje meerdere andere objecten. Als je de regel kent, is het voor zinnetjes met het woordje *pli* natuurlijk makkelijk om het goede antwoord te geven: dan kies je voor het plaatje met meerdere objecten. Voor zinnetjes met het woord *tra* is dat veel moeilijker, dan moet je heel goed hebben onthouden wat de rest van de zin ook al weer betekende. In het experiment dat wordt beschreven in dit hoofdstuk leerden 50 kleuters de kunsttaal, en wat bleek: ze gaven vaker een goed antwoord als *pli* in het testzinnetje zat dan wanneer *tra* daarin zat. Kortom, het lukte de kleuters om in een heel korte tijd zo'n grammaticale regel statistisch te leren.

Zijn kleuters zich bewust van de taal die ze leren?

De volgende vraag is of de kinderen ook doorhebben dat ze die regel hebben geleerd. Waarschijnlijk kunnen ze dat niet in mooie volzinnen uitleggen, maar misschien kunnen we wel op een andere manier achterhalen of ze doorhebben dat er regelmatigheden zijn. Alleen: hoe weten we dat kinderen zich bewust zijn van de regels in hun taal, als ze (nog) niet in staat zijn om dat te vertellen en we dat dus niet aan ze kunnen vragen? Om dat te weten te komen is het goed om eerst vast te stellen dat er meerdere vormen van bewustzijn bestaan. Op veel momenten ben je je bewust van de handeling die je uitvoert of de kennis die je hebt, en kun je dat ook mooi onder woorden brengen. Bijvoorbeeld als je weet dat je nu de samenvatting van mijn proefschrift aan het lezen bent, of wanneer je op late leeftijd Nederlands leert en uit kunt leggen dat een Nederlandse zin meestal begint met het onderwerp. Daarnaast kun je je misschien ook bewust van iets zijn zonder dat je dat exact onder woorden kan brengen. Bijvoorbeeld wanneer je doorhebt dat je wat ik nu schrijf, eigenlijk niet zo goed meer begrijpt. En je daarom dit boek maar weglegt, terwijl je niet precies kunt vertellen wat het is dat je niet begrijpt. In dat laatste geval verraadt je handelen dat bewustzijn wel. Voor kinderen geldt misschien ook wel dat ze niet onder woorden kunnen brengen welke taalregels ze leren, maar dat ze zich er wel naar gedragen. Maar: hoe weten we of dat zo is?

Om die vraag te beantwoorden heb ik me gewend tot het dierenrijk. Hoewel we als enige diersoort taal gebruiken, betekent dat niet dat we als enige bewustzijn hebben. Allerlei andere dieren hebben vermoedelijk ook door wat ze weten en wat ze doen, ook al kunnen ze dat niet verwoorden. Gelukkig is hier al allerlei onderzoek naar gedaan, bij onder andere duiven, dolfijnen, ratten en apen. Wellicht kunnen we soortgelijke onderzoeksmethodes ook op kinderen toepassen. Omdat apen nog het meest op ons lijken, ga ik daar wat dieper op in. In een experiment zagen resusapen een plaatje dat ze moesten onthouden. Later zagen ze allerlei verschillende plaatjes en moesten ze daaruit het juiste plaatje kiezen. Als ze de goede keuze maakte, kregen ze een pinda. Als ze verkeerd kozen, kregen ze niks. De apen konden ook passen en geen plaatje kiezen. In dat geval kregen ze een brokje: beter dan niks, maar minder lekker dan een pinda. De onderzoekers varieerden vervolgens de tijd tussen het eerste plaatje en het keuzemoment. Hoe langer de tijd tussen die twee momenten was, des te vaker kozen de apen voor een brokje. Daarmee gaven de apen aan zich bewust te zijn van hun herinnering aan het plaatje. Hoe langer het namelijk duurde voordat ze een plaatje moesten uitzoeken, des te groter de kans dat ze het oorspronkelijke plaatje zouden vergeten. En dat betekent: geen beloning.

Iets soortgelijks deden we in **Hoofdstuk 3** ook met kleuters. Daar probeerden we eenzelfde soort experiment als met de apen te doen met de kleuters. De kleuters hoefden geen plaatjes te onthouden en kregen ook geen pinda's. In plaats daarvan leerden ze een regel en kregen ze daar vragen over. Als ze de regel in die taal kenden, waren sommige vragen makkelijk. Er waren ook vragen die met de kennis van de regel niet goed op te lossen waren. De kinderen verdienden punten met hun antwoorden: twee voor een goed antwoord, maar geen als ze het fout deden. Ook konden ze passen en naar de volgende vraag gaan, dan kregen ze één punt. Wat nou als de kinderen doorhadden wat de regel was? Dan zouden ze passen bij vragen die met kennis van de regel niet goed op te lossen waren, en zouden ze niet passen, maar een keuze maken bij de makkelijke vragen. Hoe meer ze zich bewust waren van wat ze wisten, des te meer punten konden ze verdienen. In dit hoofdstuk leg ik uit dat deze methode die al bij allerlei dieren wordt gebruikt, ook wel eens zou kunnen worden gebruikt om er bij jonge kinderen achter te komen of zij zich bewust zijn van de regels die ze moeten leren.

In **Hoofdstuk 4** leerden we daarom 70 kleuters de eerder besproken kunsttaal en gebruikten we na afloop dit beloningssysteem om te kijken of ze zich ook bewust waren van wat ze hadden geleerd over de meervoudsregel. Aan het eind van het experiment hielpen de kinderen de hoofdpersonen van het spelletje om een rivier over te steken en weer naar huis te gaan. Om die rivier over te steken moesten die personages over een aantal stenen lopen. Als de kinderen tijdens de test een goed antwoord gaven, mochten ze twee stappen zetten, maar bij een fout antwoord moesten ze blijven staan. Ook konden kinderen passen en het volgende zinnetje horen, dan zette het personage één stap. Als een kind de regel geleerd had, zouden de zinnetjes met *pli* makkelijk en de zinnetjes met *tra* moeilijk moeten zijn. De regel die de kinderen leerden was namelijk dat wanneer *pli* in een zinnetje staat, er altijd meerdere objecten of dieren te zien zijn. In een zinnetje met *tra* kan ieder aantal objecten of dieren te zien zijn. Als de kinderen deze regel doorhadden, verwachtten we dat ze bij de *pli*-zinnen een antwoord zouden willen geven om twee stappen te zetten. Bij de *tra*-zinnen zou het juist logischer zijn als ze pasten en één stap zouden zetten, want op die zinnen was het heel moeilijk om een goed antwoord te geven. De kinderen die aan dit nieuwe experiment deelnamen leken precies dat te doen. Ze kozen er iets vaker voor om te passen bij de moeilijke zinnetjes (met *tra*), terwijl ze wel een keuze wilden maken bij de makkelijke zinnetjes (met *pli*). Oftewel: het lijkt erop dat de kleuters zich op een bepaalde manier bewust zijn van de grammaticale regel die ze hebben geleerd.

Kunnen we de kinderen dan iets uitleggen?

Dat is natuurlijk allemaal leuk en aardig, maar gebruiken de kleuters dat bewustzijn ook? Het kan natuurlijk dat ze zich bewust zijn van wat ze weten, maar daar vervolgens niets mee doen. Maar het is ook mogelijk dat het bewustzijn ervoor zorgt dat ze sneller zo'n regel oppikken. Of misschien zit het bewustzijn juist wel in de weg. Op zoek naar een antwoord op deze vragen lieten we in **Hoofdstuk 5** opnieuw een grote groep kleuters (113 kinderen) de zelfverzonnen taal leren, maar nu werd die groep in tweeën gedeeld. De ene helft kreeg een korte uitleg van de regel en werd er op die manier bewust van gemaakt. De andere helft kreeg die uitleg niet. Na de uitleg kregen de kinderen de mogelijkheid om de taal te leren. Na afloop keken we of er een effect van die uitleg was op hoe de kleuters de taal leren. Om dat leerproces te onderzoeken, testten we de kinderen dit keer op twee manieren. Zoals in de eerdere experimenten kregen ze zinnetjes te horen uit de taal en moesten ze daarbij het juiste plaatje kiezen. Daarnaast brachten we ook hun oogbewegingen in kaart terwijl ze de zinnen hoorden.

Je oogbewegingen kunnen namelijk verraden of je iets geleerd hebt: je kijkt sneller naar een plaatje dat bij een zin hoort, dan naar een plaatje dat niet bij een zin hoort. Het meten van die oogbewegingen doen we met een zogenaamde *eye-tracker*. Dat is een machine waarmee je heel nauwkeurig bijhoudt waar op een scherm je ogen naar kijken. In ons onderzoek maken we elke 8.3 milliseconde een opname van die oogbeweging: dat is wel 120 keer per seconde! Zo zien we al tijdens het horen van de testzinnen of de kinderen naar het goede plaatje gaan kijken. Het geeft dus een inzicht in hoe de kinderen de taal verwerken terwijl ze aan het luisteren zijn. Je zou bijvoorbeeld kunnen verwachten dat iemand die de regel heeft geleerd al op het eerste moment dat je kan weten wat het goede plaatje is ook naar dat plaatje kijkt. Als antwoord op de vraag of kleuters iets hebben aan de uitleg van grammaticale regels, waren we dus in twee dingen geïnteresseerd. Gaf de groep die uitleg heeft gehad meer of minder goede antwoorden? En keken de kinderen die uitleg hebben gehad eerder naar het goede plaatje dan de kinderen in de andere groep?

De kinderen in het experiment die uitleg over de regel hadden gekregen, gingen eerder naar het goede plaatje kijken dan de kinderen die geen uitleg hadden gekregen. De eerste groep deed dat namelijk al voordat het hele zinnetje was afgelopen, vanaf het allereerste moment dat je kon weten wat het goede antwoord was. De kinderen die geen uitleg kregen, gingen ook naar het goede plaatje kijken, maar pas na afloop van het zinnetje. Op basis van deze data lijkt het dus alsof er een positief effect is van de uitleg op het leren van de regel. Maar we volgden niet alleen de oogbewegingen van de kinderen. We keken ook of ze een goed antwoord gaven. En die resultaten zijn wat moeilijker te interpreteren. We hebben namelijk geen duidelijk bewijs dat de kinderen die uitleg kregen meer goede antwoorden gaven op de test achteraf. Dat is opvallend, want het laat dus iets anders zien dat wat de oogbewegingen ons vertellen.

Wat moeten we hier nu mee? Kinderen die uitleg kregen, keken sneller naar het goede plaatje, maar we hebben geen duidelijk bewijs om aan te nemen dat ze ook meer goede antwoorden kunnen geven op de test, na afloop. Betekent dat dat uitleg toch geen zin heeft? Of misschien wel maar is het effect slechts subtiel? In **Hoofdstuk 6** deden we daarom precies hetzelfde experiment nog een keer, maar nu werden de kinderen een dag later nog een keer getest. Wie weet moet de uitleg die de kinderen kregen een nachtje indalen. Bovendien luisterde de helft van de kinderen op die tweede dag opnieuw naar de zinnetjes en kregen ze zo extra taalaanbod. In dit geval testten we 103 kinderen. Opnieuw leverde het experiment niet echt bewijs dat de kinderen die uitleg kregen meer goede antwoorden gaven op de test achteraf, ook niet een dag later, maar hun oogbewegingen leken wel weer beïnvloed door de uitleg. Ook zagen we dat kinderen die op de tweede dag extra taalaanbod kregen de regel iets beter kenden.

Kortom, het is moeilijk om heel harde conclusies te trekken op basis van de experimenten die in dit boek zijn beschreven. Wat we vooral wilden laten zien is dat onderzoek naar de rol van bewustzijn wanneer kinderen taal leren, erg belangrijk is. Waarschijnlijk is het leren van een taal een complex dynamisch proces, waar statistisch leren, bewustzijn, en uitleg allemaal aan bijdragen. Hoewel de precieze verhoudingen tussen al die factoren nog onduidelijk zijn, en er nog niet veel onderzoek naar is gedaan, hoop ik de lezer er in **Hoofdstuk 7** van te overtuigen dat meer onderzoek hiernaar toch de moeite waard is.

About the author

Sybren Spit was born on February 3, 1993 in Valencia, Spain, but soon after moved to Rotterdam, the Netherlands. In 2011 he started in the BA program Dutch language and Culture at the University of Amsterdam, from which he graduated in 2014. During his BA, he developed an interest in linguistics and pursued a rMA degree in that field at the same university. At the end of 2016, Sybren won a faculty scholarship at the UvA, and started his PhD project 'Metalinguistic awareness in (early) second language acquisition'. The work from this project was presented at several conferences, is published in national and international journals, and is accumulated in this book. During his project, Sybren taught courses at the UvA on various topics, such as statistics, human cognition and philosophy of science. He also co-supervised numerous BA and MA theses. In 2019, he held a position as a lecturer in Dutch linguistics at the Open University.

Public outreach always played an important role in Sybren's work. In 2017 he became one of the Faces of Science of the Royal Academy of Sciences (KNAW) in order to popularize his work and science in general. In this role, he wrote numerous blogs for *Kennislink*, participated in *ScienceBattles*, and gave presentations at several scientific events. Sybren also held secretary and chaired *het Werkverband Amsterdamse Psycholinguïsten*, an organisation that aims to build bridges between science and practice in the field of applied linguistics.

From 2019 onwards, Sybren combined his PhD position with a position as a data analyst at the ministry of Education, Culture and Science (OCW). At OCW, he contributes to a more data-driven approach to policy making, and investigates how scientific knowledge can be used to guide development. This resulted, amongst other things, in a study on causal inference and the state exam Dutch as a second language. He also worked on a post-doc project at Leiden University, which aimed to replicate two much cited studies on language acquisition in infants.