Learning Non-Adjacent Dependencies:
A Mechanism for Language Acquisition

Published by
LOT phone: +31 30 253 6111
Trans 10
3512 JK Utrecht email: lot@uu.nl
The Netherlands http://www.lotschool.nl

Cover illustration: Andreea Manea, *Untitled*.

ISBN: 978-94-6093-236-6
NUR: 616

Learning Non-Adjacent Dependencies: A Mechanism for Language Acquisition

Het Leren van Niet-Aangrenzende Afhankelijkheden: Een Mechanisme voor Taalverwerving

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 16 juni 2017
des middags 4.15 uur

door

Ileana Cristina Grama

geboren op 15 September 1987 te Boekarest, Roemenië

Promotoren: Prof.dr. F.N.K. Wijnen
Prof.dr. P.H.A. Coopmans

This research has been supported by the Netherlands Organization for Scientific Research (NWO).

In loving memory of my mother and grandmother

Table of Contents

Acknowledgments xi
Tables and Figures xix

CHAPTER 1
Introduction 1

1.1 Introduction 1
1.2 Non-Adjacent Dependency-Learning 6
1.3 NAD-Learning as Rule-Learning or Statistical Learning 8
1.4 The Nature of NAD-Learning 10
1.5 NAD-Learning in Development 12
1.6 Dependencies in Natural Language 15
1.7 Learning Morpho-Syntactic Dependencies in Natural Languages 18
1.8 Purpose of the Current Study 22

CHAPTER 2
The Role of Prosodic Cues in Non-Adjacent Dependency Learning 25

2.1 Perceptual Cues to Functionam Morphemes 26
2.2 Functors in Early Acquisition 28
2.3 The Role of Perceptual Cues in NAD-Learning 30
2.4 Purpose of the Current Study 32
2.5 Research Question and Implementation 35
2.6 Experiment 1 36
    2.6.1 Participants 38
    2.6.2 Materials 38
    2.6.3 Procedure 41
    2.6.4 Results 42
    2.6.5 Discussion 44
2.7 Experiment 2 46
    2.7.1 Participants 46
    2.7.2 Materials 46
    2.7.3 Results 46
    2.7.4 Discussion 48
2.8 Experiment 3 51
    2.8.1 Stimuli 51

2.8.2 Participants 51
2.8.3 Results 51
2.8.4 Discussion 52
2.9 General Discussion 54
2.9.1 NAD-Learning Guided by Gestalt Principles of Perception 55
2.9.2 Caveats and Questions for Future Research 57

CHAPTER 3
Incremental Learning of Non-Adjacent Dependencies 61

3.1 The ‘Starting Small’ Hypothesis 62
3.2 Reversing the Argument: The ‘Starting Big’ Hypothesis 68
3.3 Starting Small and Learning Dependencies 70
3.4 Research Questions 72
3.5 Experiment 1: Varying the Length of Pre-Familiarization 74
3.5.1 Materials 74
3.5.2 Participants 75
3.5.3 Procedure 75
3.5.4 Results 76
3.5.5 Discussion 77
3.6 Experiment 2: Emphasized Short Pre-Familiarization 78
3.6.1 Materials 78
3.6.2 Participants 79
3.6.3 Results 79
3.6.4 Discussion 80
3.7 Experiment 3: Simple Pre-Familiarization 82
3.7.1 Materials 82
3.7.2 Participants 83
3.7.3 Results 83
3.7.4 Discussion 83
3.8 Experiment 4: Controlling for Pre-Familiarization and Instruction 87
3.5.1 Materials 88
3.5.2 Participants 88
3.5.3 Procedure 88
3.5.4 Results 89
3.5.5 Discussion 89
3.9 General Discussion 90

CHAPTER 4
Generalization and Prosodic Cues to Non-Adjacent Dependency-Learning in Infants….. 97

4.1 Experiment 1 99
4.1.1 Introduction 99
4.1.2 Materials 102
4.1.3 Participants 103
4.1.4 Procedure 104
4.1.5 Results 106
4.1.6 Discussion 109
4.2 Experiment 2 113
4.2.1 Introduction 113
4.2.2 Materials 114
4.2.3 Participants 115
4.2.4 Procedure 115
4.2.5 Results 116
4.2.6 Discussion 118
4.3 General Discussion 121

CHAPTER 5
Discussion and Conclusions 125

5.1 The Role of Prosody in NAD-Learning 126
5.2 The Role of Incremental Learning in NAD-Learning 129
5.3 Generalizing NADs to Novel Contexts 132
5.4 NAD-Learning as a Mechanism for Language Acquisition 134
5.5. What We Have Learned About NAD-Learning 137
5.6. Methods and Questions for Future Research 141

Bibliography 145
Appendix 161
Tables for Chapter 2 161
Tables for Chapter 4 165
Samenvatting in het Nederlands 169

Acknowledgments

"...It's not the department, it's a subject, like Mechanical Avunculogratulation or Pylocatabasis. They all fall under the same heading of Tetrapyloctomy."

"What's tetra...?" I asked.

"The art of splitting hairs four ways. This is the department of useless techniques. Mechanical Avunculogratulation, for example, is how to build machines for greeting uncles.

(Umberto Eco, *Foucault's Pendulum*)

The acknowledgements section of any PhD dissertation is most often a testimony to the fact that the author used to (and/or still secretely does) nourish ambitions of being a writer. I'll come clean now and admit that I was definitely aiming for artistry more than function when I first sat down to write this section.

Until I realized the *number* of people I actually have to thank in here. It's astronomical. So I do not have time to parade any originality or writerly sense of humour – instead I'll be moderately sentimental and unoriginally nostalgic about a number of years I have spent in the company of some great people. Ok, here goes.

Thin though it may appear, this ode to Tetrapyloctomy that is my dissertation has been the result of a considerable amount of effort from people much kinder and smarter than me. The first of them is Frank Wijnen. Frank has had the good nature of tolerating my wild psycholinguistic goose-chases since I was a wee master's student running my first experiment in the phonetics lab, and has always encouraged me. He has patiently guided me through the barren landscape of resultless experiments all the way through to a small oasis of (some) statistical significance. We have talked about my research for many more hours than would have been wise to dedicate to a measley PhD student when he had so many other responsibilities. He's been a real mentor, with all that that entails. Thank you, Frank. I mean really, *thank you*!

Equally worthy of thanks are Annemarie Kerkhoff and Peter Coopmans. Annemarie has fulfilled the part of a supervisor for much of this PhD, despite not being credited for it. She has met with me, talked to me, agreed with me, disagreed with me, shared *her* data with me, looked at *my* data with me, perused my tangled writing with a fine comb of common sense and logic and much, much more. I am

hugely indebted to you, Annemarie, for your patience and guidance, and your attention to detail which really is a rare and necessary quality for an experimenter. Thank you Peter for the patience with which you read my work and discussed it with me, for making my writing good and my arguments clear, and my thesis readable. Also for helping and guiding me through designing and teaching a course entirely by myself, which I would never have had the courage to do without your encouragement. Your support has been invaluable.

I am also grateful to Sergey Avrutin, who has not only given me his thoughts on my research, but has also co-opted me as a teacher in his courses time and time again. I have greatly enjoyed teaching with you Sergey, and I hope to continue to do so at least for a little while longer. Along with René Kager, Sergey has also been my go-to authority when I needed a fresh perspective on my research – I thank them both for taking the time to discuss my experiments, and for steering me in the right direction and asking the right questions.

I thank my reading committee for agreeing to read my dissertation and be involved in my defense: René Kager, Sergey Avrutin, Hugo Quené, Claartje Levelt and Ruth de Diego-Balaguer. Special thanks to René and Hugo for their detailed comments, which have allowed me to sharpen some of my ideas or consider some arguments more in depth.

I am particularly indebted to Ruth de Diego-Balaguer, who has been kind enough to invite me to Barcelona to discuss my research, and also to Anna Martìnez Àlvarez – I greatly enjoyed my time in Barcelona with them, as well as their thoughts, ideas and comments. Thank you, Ruth, for agreeing to sponsor some so far unsuccessful grant applications, and for taking an interest in my ever-evolving research proposals; also thank you for agreeing to mangle your June conference schedule to be there at my defense – I really look forward to having you there!

While I am still on the topic of mentorship, I cannot go without mentioning the people that have been a constant help to any PhD student of my generation: Martin Everaert and Maaike Schoorlemmer. Thank you, Martin, for the interest you take in everyone's research no matter how distant (or unchomskian) from your own. Thank you for teaching with me, getting me into super-exclusive events like the 2015 KNAW Colloquim 'The Biology of Language', for sending me to Kosiçe for an Erasmus teaching exchange. Thank you for all the amazing opportunities you have given me. And speaking of Kosiçe, thank you also to eveyone at the Pavol Jozef Šafárik University who has made my teaching stints there a real pleasure: Pavel Stekauer and Lívia Körtvélyessy, Silvia Rosivalová Baučeková (thank you for our lunches and our rants against the establishment ☺), Lýdia Borková (thanks for trekking the best parts of Kosiçe with me) and everyone else.

Thanks to Maaike also for her constant support, she has been my go-to person for any questions or problems since day one of my setting foot in Utrecht as a master's student. She's basically the first person I ever contacted at UU, back in a time where my knowlwedge of Dutch was so limited that I thought Maaike is read as 'Mike' (and not 'Maai - kuh') and is, in fact, a man's name. And she was so much nicer than anyone would be if they received an email starting with ''Dear Mr. Schoorlemmer'.

I also owe a great debt of gratitude to Iris Mulders and Maartje de Klerk, without whom I could not have done any of the work in this dissertation. Iris has taught me everything I know about running an experiment, has helped me set up all of the adult experiments, recruited all of my participants, and patiently sufferred through all the promises I made that this would be my last experiment. None of them came true. Maartje has done the same for my infant experiments, helping me set them up and run them, discussing the results with me and giving me constant advice.

Last but not least on this list of people I look up to as mentors: Eddy Ruys & Riny Huijbregts. Eddy was my teacher for one of my first courses in Utrecht (Syntax), and quickly became my role-model. He gave me my first gig as a teaching assistant, and it was the most fun I ever had teaching. Riny was *the* first teacher I had in Utrecht, at a summer school long ago: his passion in teaching us the most interesting things I had never heard of in linguistics convinced me that I had to come to Utrecht for an MA. Riny, says Eddy, is the reason Eddy went into linguistics. Riny and Eddy are the reason I went into linguistics.

And now to my colleagues and friends.

First and foremost I have to acknowledge the contributions of people who have worked on collecting and analyzing the data in this dissertation. I thank Silvia Radulescu for running Experiment 2 in Chapter 2, and the conversations we've had ever since about our experiments, our predictions, the academia, etc. Silvia's doing some exciting research at the moment and I for am very curious to see her results!

I thank my wonderful interns who have done a lot of great work: Lianne van Setten, Fieneke Jochemsem and Suzanne Dekker for running a large part of the infant experiments in Chapter 4. Stella Schram for running some adult experiments that, though inconclusive, gave me much food for thought. I also thank Sule Kurtcebe, Cora Pots and Karlijn Kouwer for their help in running the infant experiments. I am also grateful to Kirsten Schutter for countless discussions and directions on how to statistically analyze my data, and to Roel Willems for his advice on analyzing the adult data of Chapter 2.

Secondly, I want to thank my lovely paranymphs, Marko Hladnik and Katrien van Riet. They have been my friends for a long time, and feature in some of my best memories that I have as a PhD student. Thank you Marko and Sirin for adopting me to your boardgame afternoons a long time ago; for the movie nights, the escape rooms, the dinners, and the GoT exhibition where I met this guy who got run over by a car. Thank you Katrien for letting Ben and me live in your wonderful house despite making all the neighbours think that we are Mormons; thank you for your risottos, the boardgames, the running, that 40km walk, and a bunch of other things. You are the best sister-wife. I could not think of better godparents than Marko and Katrien for this dissertation, and could not imagine walking into my defense without them at my side.

I'm happy to have been part of a lovely group of people who met every Friday to talk about their experiments, puzzle over statistics, complain about restless babies or listen to the sound of strange made-up words. Thanks to everyone who has been a part of the babylab meetings: Maartje, Frank, Silvia; Liquan Liu, my BUCLD companion in 2011, who is sorely missed for his fabulous sprints on the corridors of Janskerkhof 13; Brigitta Keij, my teaching buddy for many years, who graduates a few months before me to show me how a pro does it; Carolien van den Hazelkamp, my fellow PhD-er with as great a flair for methodological nitty-gritty stuff as her sense of humour; Chen Ao, also my BUCLD companion, who makes giant, awesome research projects seem easy to handle; Sule Kurtcebe, who knew all the ins and outs of this lab and is also sorely missed; Desirée Capel, my one-time internship supervisor and longtime co-sufferrer in matters of statistical analysis and the pitfall of artificial grammar learning research; Lisanne Geurts, Mengru Han, Emma Everaert, Jessica van Schagen, Niloofar Hashemzadeh, Shanshan Fan, Herbie Liu, and everyone who, over the years, have made these meetings the highlight of my week, and a source of good advice as well as good humour.

I'm also thankful for the other groups and gatherings I've been part of these past few years: EMLAR – thanks to Shalom Zuckerman with whom I've spent many an hour poring over the EMLAR website, as well as Sergio Baauw, Maaike, Aoju Chen, Hayo Terband, Elena Tribushinina, Hans Rutger Bosker, Anna Sara Romoren and Sandrien van Ommen; Uiltjesdagen – thanks to Marjolein van Egmond and Eva Poortman for all the fun we had posting clues around town for the treasure hunt; and ELiTU – thanks to Sandrien, Stavroula Alexandropoulou, Anna Sara, Anne-France Pinget, Mengru and Sophia Manika for all those lovely Monday afternoons with cookies and tea and experimental linguistics.

And I'm very thankful for all the other gatherings that have made these PhD years feel much sunnier than Dutch weather would realistically allow: to Mirjam Hachem for all the coffee and the long talks about books, the academia, feminism and biolinguistics, not necessarily in that order; Bert LeBruyn for still being my role model, though sadly no longer my office mate; Marko Hladnik, Sirin Tugbay, Heidi Klockmann and Lukasz Glab for the boardgames, escape rooms, ping-poing matches and áatacks ☺; Maartje Schulpen for being a lovely (and very patient!) housemate; Anna Volkova, Liv Persson and Anna Sara for a wonderful biking trip somewhere in 2012 that remains the best biking trip I have ever attempted; Marko Simonovic for being the zany office mate who always walks in exactly in the few seconds a day that I am checking my Facebook; Anne van Leeuwen, who also used to be me office mate but never saw how much I was on Facebook ☺; Lotte Dros-Hendriks, for taking a particular interest in this humble thesis; Anja Goldschmidt, Will Schuerman, Björn 't Hart, Jolien Scholten, Franca Wesseling, Dominique Blok, Myrthre Bergstra, Marta Castella, Loes Koring, Koen Sebregts, Marijana Marelj, Alexis Dimitriadis, Jakub Dotlacil, Yaron McNabb, Alexia Guerra Rivera, Assaf Toledo, Zenghui Liu, Anqi Yang, Jet Hoek, Suzanne Kleijn, Yipu Wei, Erin Pretorius, Nadya Goldberg, Marijke de Belder, as well as the above-mentioned Mirjam, Bert, Lotte, Marko H., Marko S., Heidi, Brigitta, Liquan, Ao, Carolien, Silvia, Liv, Anna V., Anna Sara, Anne-France, Anne, Maartje, Stavroula, Marjolein, Eva, Sandrien, and all those with whom I have shared offices, lunches, picnics, Friday afternoon drinks, impostor syndrome, general deadline-related stress, outrageous teaching stories, sunny afternoons in the courtyard or rainy mid-days in the cafeteria.

Many more people, outside of my institute, have made my PhD years the best I could've had. Thank you to my very first flatmates Dani and Carlos, who made Utrecht fun for me when I first moved in. Thank you to my lovely litkids for the litkid dinners (the main reason I ever learned to cook), trips to Vlieland, moving house, and all the great memories we've shared: Katrien, my above-mentioned sister-wife; Joke Brasser and Joris Hiemstra, our foster parents and great running buddies (and Leonie Brasser too); Peter Maurits, you beautiful stress-bunny, if anyone deserves to have it all in the academia, it's you; Paul Avery, there is no describing your awesomeness other than we all want to adopt you; Abby Waysdorf, the Eurovision song contest just isn't the same without you; Roni Dragomir you are sorely missed as the perfect coffee date, stay-over host and guide to Eindhoven; Maik Arets, wish you still lived next door and dropped by to watch Dylan Moran; Enno Maessen, lil'E you are still the ultimate fratboy, the pointiest shoes, the shiniest coiffe; and Casper van de Veen, big C, for you I save the best word in the dictionary: *genieten*. Thank you Maria Zirra and Sebastian Bican for being my homegirl/boy in this crazy town: I'm honored to have been on your superhero team from highschool to MA, with all the cooking,

eating, drinking, dancing, travelling, working, being lazy, reading, movie-watching, navel-gazing, time-wasting and shit-talking that that entails (and if we're talking superhero team here, Marius Motoca as the fourth Horseman of the Apocalypse who roamed the halls of Uni with us).

Thank you also to the awesome Law people who have taken me under their wing: Raluca Racasanu and Andreea Manea, friends that I have stolen from Ben and now they are mine ☺. I'm particularly grateful to Andreea for making the beatiful cover illustration of this book. Arron Honniball, thank you for some really crazy trips like the Rotterdam container port (which no-one in their right minds would have predicted to be so much fun). Jurriaan de Jong thank you for putting up with my less-than-gracious wins at Settlers of Catan. Caio, thank you for your lovely Brazilian food, you are sorely missed at Sunday boardgames. Thank you Lucas Brouwers and Dani Heerdt, for mid-summer barbecues and running slowly enough for me to catch up. Thank you Marize Verhagen for awesome dancing, thank you Hanne Cuyckens for your sense of humour after about 30km of walking, thank you Hester Kroeze, Mariette Hiemstra and Otto Spijkers for participating in my experiments.

Special thanks to my Romanian mafia all around the place: Cristina Dumitrache, our protest- and rowing-buddy; Bogdan Ivanel and Maria Neicu, as well as Marc Stefanescu and Marloes, for dinners, beers, cocktails and general posh fun; Diana Popescu & Alex Marcoci, Alexe Zirra and Trotsky (Andrei Sorescu), Irina Mnohoghitnei & Andrei Troie, for a great stay in London; Dragos Costache, Ileana Zirra, Mihnea Florea, Ana Goanta, Ruxi Marcu, Elena Burlacu, Sofia Costescu, my childhood friends Ali & Denisa, and all the people who made coming back to Bucharest so much fun!

Thank you Ana Turos, my one-time deskmate, my all-time New Year's Eve dancing buddy, my favourite actress in the whole wide world – I've been so lucky to be your friend for so many years, thank you and your mum, Anca Hariton, for always sharing the best booze and gossip with me!

A special thanks goes to my teacher, Atena Serban, who taught me so much more than just English over the years. Thank you for being my guardian angel, my wise advisor, thank you for believing in me and encouraging me!

Finally thank you to the Thompson family, Jonny, Lidia, Simon, Daniel, and Babcia, for welcoming me into your family and your home, for all our trips, for being the most wonderful, fun, kind, warm, and grounded British family I know!

I wouldn't be anywhere in the world today without the support of my family, who has overcome so many adversities that I would've never, in my best dreams, imagined we'd arrive where we are today.

If I am here at all it is mainly thanks to the efforts of my amazing aunt Sanda, who comes from a long line of strong brilliant women who carry the world on their shoulders. As a scientist and a woman she has always been an inspiration to me. I thank her for challenging me to come to Utrecht and pursue a PhD, for being my pivot in an unfamiliar land when I first came here, for her unwavering support, her patient advice, her help with things both small and big. I thank her for taking the heaviest burderns, being our anchor through the most difficult times, for always keeping close. I hope that all your hard work, your ethics, your dedication, your caring for the people around you, your sensibility, your sense of humour, your wisdom, your compassion and your selflessness return to you a million times over!

I've never had siblings but I've always been lucky to have a cousin that was better than any sister I could've had. Thank you Dora for easily forgetting how annoying I was when you were small and how much I teased you. I was a douchebag but I was still thrilled when you came over and I still missed you like hell everytime you had to go. Thank you for sharing all your boy stuff, your friend stuff or your school stuff with me, despite all the moralizing that I do. Thank you for always calling me up, for spending hours gossiping with me, for your evil sense of humour. I'm really impressed with you, you know, you're pretty brilliant and it's no wonder you've got so many people who love you and stick up for you.

For a little Romanian interlude: multumesc bunicilor mei, mama Gabi si tata Liță, care mi-au fost alaturi dintotdeauna, care au avut grija de mine, s-au laudat cu mine si au crezut in viitorul meu. Sper ca v-am facut mandri de mine si va dedic aceasta teza voua, in amintirea mamei mele, Dana, si a matusii Miți.

Ii multumesc tatalui meu Serban, pentru intelepciunea si rabdarea lui, pentru discutiile noastre serioase si mai putin serioase, pentru excursiile noastre la munte, pentru filmele proaste si filmele bune, pentru jocurile de Scrabble la care ma bate de obicei, pentru plimbarile prin parc si toate lucrurile pentru care de-abia astept sa vin la Bucuresti in fiecare luna. Nu stiu daca iti dai seama cat de multe am invatat de fapt de la tine, si cat de mult a insemnat pentru mine sa te am alaturi in toti acesti ani. Viata n-a fost intotdeauna cea mai usoara, dar sper ca de acum inainte sa avem numai zile senine!

Si mama. Bunica mea careia intotdeauna i-am zis mama. Iti multumesc din suflet ca ai scos familia noastra la liman din cele mai grele incercari. Fara tine nu am fi supravietuit niciunul. De la tine am invatat sa iubesc cartile, muzica, muzeele, parcurile, visinele vara, Sucevita, pianul, Palatul Copiilor, cititul pe o

banca, si opera de la radio seara. Inca te mai vad pe banca la Radauti langa mine, cu o carte in mana, sau inmanandu-mi prima brosa cand ne pregateam sa ma duci la teatru de papusi. Aproape ca am uitat cat de greu ti-a fost in toti acei ani in care te-ai luptat cu bolile (si ale noastre si ale tale), cu lipsurile, si cu toate greutatile, fara sa te lasi. Dar inca te vad jucand table cu mine, si inca ma trezesc diminetile de Anul Nou asteptand sa vii sa ne uitam impreuna la concertul de la Viena. Sa stii ca ne e tare dor de tine, tuturor, si speram ca acolo unde esti te-ai lepadat de toata suferinta omeneasca. Si ca esti cat de cat mandra de noi, ca numai prin sacrificiul tau am ajuns acolo unde suntem astazi. Eu una sper sa pot sa fiu vreodata macar un sfert din omul care ai fost tu.

There is just one more person I need to thank for being with me all these years. My dearest, closest, most beloved Ben. Otherwise known, in small, informal circles of friends, as Sugartits. Ben is the Dale Cooper to my Harry S. Truman, not only handsome and wise, but also very preoccupied with the plight of the Tibetan people. He is the Roy to my Moss, the Calvin to my Hobbes, the Brian to my Marsha, the Vince Noir to my Howard Moon, the Jack Donaghy to my Liz Lemon, the Dave Lister to my Arnold Rimmer. Ben, I am very lucky that you did not suffer a gruesome death in some car accident or nightclub altercation before I got to meet you. And I will earnestly endeavour to ensure that you will not do so in future. Because I would like to have you around to make fun of me for a very long time. Ben, you're a positive force in the universe, and not just thanks to your awesome dancing moves, or because you save a lot of African people from the jaws of evil transnational businesses. But because everyone's life is better when you're around. You've given me more confidence, you've talked me through the hardest parts of life, you've made me be less of a sore loser at boardgames, you've taught me to judge less, drink more, row faster, eat less healthy, and cheat at Scrabble. You've introduced me to all the best British comedy, and I have introduced you to ice tea and pesto, so we're even. Honestly, B.J., you're my thunder-buddy for life!

Tables and Figures

Nr.	Table	Page
C2.1	Results of Experiments 1-2 (Chapter 2)	43
C3.1	Results of Experiment 1 (Chapter 3)	76
C3.2	Results of Experiment 2 (Chapter 3)	80
C3.3	Results of Experiment 4 (Chapter 3)	89
C4.1	Results of Experiment 1 (Chapter 4)	107
C4.2	Results of Experiment 2 (Chapter 4)	116
A2.1	Stimuli *a/b* used in Experiment 1 (Chapter 2)	161
A2.2	Stimuli *X* used in Experiments 1, 3 (Chapter 2)	162
A2.3	Stimuli *a/b* used in Experiment 2 (Chapter 2)	163
A4.1	Stimuli *a/b/X* used in Experiment 1 (Chapter 4)	165-6
A4.2	Stimuli *a/b* used in Experiment 2 (Chapter 4)	167
	Figure	
C2.1	Results of Experiments 1-2 (Chapter 2)	42
C4.1	Results for Experiment 2 (Chapter 4)	117
A2.1	Pitch contour of *aXb* in Functional Condition(s) (Chapter 2)	164

Chapter 1
Introduction

1.1 Introduction

The present study centers on a core notion that is as familiar and ubiquitous as it is difficult to define or understand: learning. Loosely defined, learning can be construed as a process by which a learner exposed to input (of whatever kind) develops mental representations consistent with that input that may inform their future behavior. In the case of linguistic input, linguists have strived for decades to understand the mechanisms by which a child learner exposed to utterances in any natural language develops, within the first few years of their life, mental representations that capture the full complexity of that language, enabling them to subsequently use these representations to produce and understand language in real time.

To study learning as it is happening (in the laboratory), researchers must seek to dissociate it from knowledge or biases that subjects may exhibit based on prior experience, otherwise it is unclear whether the behavior of the participant reflects only learning or a combination of learning and prior knowledge. A method that is particularly appropriate for this purpose is Artificial Language Learning (ALL). Subjects, whether adults, infants, or even non-human species, are exposed to stimuli – linguistic or non-linguistic, visual, auditory or even tactile – in a controlled environment, for a limited amount of time. These stimuli are combined following certain rules, having an underlying pattern that learners are not informed of. Because the exposure material is created by the researcher for the purpose of the experiment, it is novel to the subject, decreasing the likelihood of interference from prior experience[1]. Furthermore, the researcher can limit the type and amount of information present in the input, trying to ensure that only a specific type of

[1] Note, however, that although learners of an artificial language have no prior knowledge of the particular language they are learning, they may still be using their prior experience of natural languages in an ALL task. Namely, they may be drawing analogies between the artificial language they are learning and the natural language(s) they speak, or their native language may bias them to look for specific patterns in an artificial language. Few ALL studies directly address the possibility of transfer between one's native language and the artificial language to be learned; in this thesis I explicitly discuss the possibility of transfer whenever it becomes likely that the participants' linguistic experience may be guiding their learning of an artificial grammar.

mental computation can be successfully applied to discover the structure of the input. Following the exposure, subjects are tested on what they have learned in a variety of ways: they are asked to discriminate or rate (new) correct or incorrect patterns according to the AL presented, or reproduce those patterns from memory, or react to in/correct patterns in real time, etc.

Arthur Reber's seminal work (Reber 1967, 1989) showed that adults exposed to a set of letter strings (e.g. *ATXTBBBG*) generated by a finite-state grammar (a set of rules that govern the transition from one letter to the next) were not aware of the grammar rules explicitly, but showed sensitivity to them implicitly by preferring strings that conformed to the rules. Thus, Reber (1967) presented learners with such a grammar and asked them to memorize the string sets that they were presented with. He showed that subjects could subsequently reproduce strings that conformed to the grammar far better than strings that didn't, and that with prolonged exposure their accuracy in memorizing and reproducing grammatical strings improved over time. Other research (Reber & Allen, 1978) showed that memorization was not necessary, and that simply attending to the input without completing any task lead to more robust learning (measured as the rate at which learners accepted strings generated by the grammar and rejected strings that violated the grammar rules at test).

Reber's work opened up several avenues for research: firstly, it led to studies further investigating cognitive aspects of human ability for pattern-learning from mere exposure, its limits, its connections with other cognitive faculties (e.g. memory, attention), and the nature of the representations it produced. Thus, it was shown that (implicit) learning in ALL experiments may yield long-term effects, with the patterns of the artificial grammar being preserved in long-term memory for as much as two years (Allen & Reber, 1980), that learners may develop abstract representations of the patterns they hear, as they are able to transfer those patterns to a novel (e.g. *FEDELLLM*) set of stimuli (Gómez & Schvanenveldt, 1994), and that implicit (as opposed to explicit) learning is possible even in patients with severe impairments of declarative memory (amnesia; Reber & Squire, 1998). Implicit learning, defined as learning in the absence of awareness (see Cleeremans, Destrebecqz & Boyer, 1998 and references therein) emerged as a specific type of learning (of patterns or regularities in the input) that required minimal effort and produced robust and long-lasting effects .

A second direction of study employed computational models to simulate the learning that adults showed themselves capable of in behavioral experiments (Cleeremans & McClelland, 1991; Cleeremans, 1993; Perruchet, Vinter, Pacton & Gallego, 2002; Boucher & Dienes, 2003). Finally, a third direction employed the ALL paradigm to investigate new types of patterns that subjects (adults, infants, etc.) could detect effortlessly from input and to connect this pattern-detection ability to language acquisition. It is the latter that I turn to in the remainder of this chapter.

Early ALL studies showed that both infants and adults were capable of exploiting transitional probabilities of syllables (the probability that syllable B follows syllable A, calculated as the ratio between the frequency of the AB sequence and the frequency of either of either of the A or B syllables alone) to segment word-like units from continuous input (Saffran, Newport & Aslin, 1996; Aslin, Saffran & Newport, 1998). This type of 'statistical learning' seemed particularly relevant to language acquisition, as spoken language offers no consistent pause cues to the boundaries of words. If a phrase like *prettybaby* is presented to a listener with absolutely no knowledge of English, this person would have no straightforward cue to identify the words *pretty* and *baby*. However, with more input, the learner would hear the syllable *–tty* following the syllable pre- more consistently, since *pretty* is a word of the language; the sequence *tty-ba* would have a lower transitional probability as it is not a word in the language, and the syllable *–tty* could precede a large variety of other syllables (e.g. *pre-tty flo-wer*, *pre-tty girl*, *pre-tty doll*, etc.). Further research pitted this learning mechanism against other cues to word segmentation (word stress, Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003), or showed that it failed to produce a significant learning effect with more complex input (Johnson & Tyler, 2010). It is as yet difficult to determine how big a role statistical learning of transitional probabilities may actually play in language acquisition; for now, the finding that learners can effortlessly exploit statistical information in the input provides evidence for the existence of statistical learning mechanisms, and such mechanisms could conceivably subserve language acquisition.

Subsequent ALL studies showed that humans were capable of exploiting not only transitional probabilities between syllables, but a variety of other statistical information from the input. Thus learners (both adults and infants) could exploit the distribution of acoustic properties (of speech sounds) in order to determine which sounds were realizations of the same underlying phonemic category or which sounds represented distinct phonemic categories (Maye & Gerken, 2000; Maye, Werker & Gerken, 2002). This ability was hypothesized to aid young infants in detecting the phonological categories of their native language.

Learners were also shown to be sensitive to the similarities between words that occurred in the same context (or 'frame') in the input (Mintz 2002, 2006). This was hypothesized to aid in the formation of morphological categories (noun, verb, etc.) as members of the same morphological categories often occurred in the same 'frequent frames' (e.g. *to_it* is a frame that frequently occurs in English, and usually combines with an intervening verb, e.g. *to leave it*, *to miss it*, etc., cf.

Mintz, 2003). Finally, learners were shown to detect high transitional probabilities not only between adjacent, but also between non-adjacent syllables, phonemes, or other linguistic units (Gómez, 2002; Newport & Aslin, 2004), an ability which could support, as I will argue in this dissertation, the acquisition of morpho-syntactic dependencies in natural languages (e.g. the dependency between the auxiliary *'is'* and the progressive suffix *'ing'* as in *The cook **is** knead**ing** the bread*, see Santelmann & Jusczyk, 1998).

What these studies suggest is that humans (adult of infant) are capable of a variety of mental computations (category-formation, distributional analysis or combinatorial statistics) and these mental computations could, in principle, be applied to linguistic input to glean important properties of natural languages. What is more, some of these learning mechanisms have been shown to extend beyond the linguistic domain: for instance, humans have the ability to exploit transitional probabilities not only between linguistic units, but also between visual stimuli (Fiser & Aslin, 2002a & 2002b; Kirkham, Slemmer & Johnson, 2002; Bulf, Johnson & Valenza, 2011) or between musical tones (Saffran, Johnson, Aslin & Newport, 1999; Creel, Newport & Aslin, 2004). Furthermore, statistical learning is not unique to the human species: cotton-top tamarins have been shown to segment groups of syllables from continuous strings based on transitional probabilities (Hauser, Newport & Aslin, 2001), and even to track co-occurrence probabilities between non-adjacent units (Newport, Hauser, Spaepen & Aslin, 2004).

It is important to note, however, that language is a uniquely human mental capacity, and that its acquisition has been argued to rely more on innate, language-specific representations than on domain-general learning mechanisms that detect surface properties of the input (Chomsky, 1980). If (some of) the learning mechanisms reviewed above are domain-general and shared with other species, their role in language acquisition must be limited, since they do not seem to capture the clear superiority humans seem to have over non-human species in acquiring language. Furthermore, these learning abilities are attested with adults as well as infants, with few differences between age groups emerging (though see Hudson Kam & Newport, 2005, 2009; Austin & Newport, 2011). Nevertheless, differences between age groups do exist in language acquisition, with younger learners showing superior language learning abilities to older ones (Newport, 1988, 1990).

Finally, a compelling argument put forth by Chomsky against the role of distributional learning mechanisms is the fact that examining surface, distributional properties of language input may not lead to the correct generalizations. Thus, hearing sentences like *The man is happy* and *Is the man happy?* might lead to a variety of potential inferences about question formation (front the third word, front *is*, front the first verb, etc.) that prove incorrect when transforming a more complex sentence like *The man who is tall is happy* into a question. To derive the correct question *Is the man who is tall happy?* the infant must learn that the fronted element is the auxiliary verb of the main clause, thus capturing the hierarchical representation for language (a subordinate clause is embedded in a main clause) that is arguably not detectable if we apply distributional learning mechanisms to the input in its linear form. Instead, Hauser, Chomsky & Fitch (2002) proposed that the special ability unique to humans that makes language-learning possible, and facilitates the acquisition of question formation, is their ability to detect recursive structure (structure that can hierarchically embed a constituent within a larger consistent).

In this dissertation I do not take a position with respect to innateness or the importance of recursion to language; instead I argue that these questions are orthogonal to the properties of distributional learning mechanisms and their role in language acquisition. While the ability to detect recursive structures may aid learners in detecting the grammatical structures that are possible in any language (as all languages arguably exhibit hierarchical structure), other mechanisms like statistical or distributional learning, category-induction, etc. may help identify properties that are specific to a language – such as segmenting specific words, identifying their combinatorial (syntactic) properties, etc. Thus, learners may need recursion to identify the syntactic structure of a sentence like *Is the man who is tall happy?* in any language, but before doing this they will require distributional learning mechanisms to segment the individual words of the sentence, establish the fact that *man* is a noun and *tall* and *happy* are adjectives or conclude based on the combinatorial properties of *the* that it is a determiner. I propose therefore, that distributional learning mechanisms are important tools through which learners may detect the units and idiosyncratic properties of the specific language they are acquiring; even if certain representations are innately specified, it will be necessary to map the input onto such representations, since the input does not come with boundaries between linguistic units or labels that identify them.

To conclude this section, humans possess the ability to spontaneously detect a variety of structures in the input. These learning processes and their features can be studied in a controlled environment in ALL studies, and might be recruited in various functions such as language acquisition, particularly in detecting structures which cannot be innately pre-specified, but must be learned from the input. Therefore, it is worthwhile to study structure-detecting learning mechanisms and to investigate what role these mechanisms could play in language acquisition. In the following two sections I focus on a specific learning mechanism (the ability to detect non-adjacent dependencies), review studies that have investigated its properties (Sections 1.2 and 1.3), proposals about the nature of this mechanism (Section 1.4) and its developmental trajectory (Section 1.5). I discuss the type of structures in natural languages that could be detected using such a mechanism (Sections 1.6 and 1.7). I set up the aim of this dissertation (Section 1.8) as providing evidence that the properties of this learning mechanism make it a potentially reliable tool for language acquisition.

1.2 Non-Adjacent Dependency-Learning

The topic of the current dissertation is a learning ability that allows both adults and infants to identify co-occurrence patterns between elements that are non-adjacent in a sequence. Thus, in a string of elements of the type ABCDEF learners will be sensitive not only to the presence of co-occurrence patterns between adjacent elements like AB or DE, but might also detect if for instance the occurrence of B consistently predicts the occurrence of E over two intervening elements. This would, for instance, prompt learners to subsequently rate a string like XBYZEW as well-formed according to the grammar, and a string like XBYZFW as ill-formed. Following initial studies showing that humans are sensitive to transitional probabilities between adjacent units in a string (Saffran, Newport & Aslin, 1996; Aslin, Saffran & Newport, 1998), the question arose whether learners are sensitive to co-occurrence statistics between non-adjacent elements as well, since natural languages clearly also exhibit co-occurrence patterns between non-adjacent elements (e.g. *The princess **is** gently kiss**ing** the frog*, or *The princess **has** repeatedly kiss**ed** the frog to no avail*).

To verify the extent to which learners can develop a sensitivity to *non-adjacent dependencies* (NADs), ALL studies present learners with strings of nonce words with the structure a_iXb_i, where the form of a_i predicts the form of b_i with high (often 100%) probability whatever the form of the intervening X. In a typical set-up, subjects listen to the unfamiliar language for a given (limited) amount of time; subsequently, they are exposed to test stimuli either respecting or violating the non-adjacent regularities heard during familiarization. Learning is evidenced as the participants' preference for correct a_iXb_i dependencies either over incorrect a_iXb_j dependencies (where the final element is not predicted by the first), or over scrambled strings like $X\,b_i\,a_j$ (Gómez, 2002; Peña et al., 2002; Onnis, Monaghan, Christiansen & Chater, 2004; Newport & Aslin, 2004; Gómez & Maye, 2005; Endress & Bonatti, 2006; Endress & Mehler, 2009; van den Bos, Christiansen & Mysiak, 2012, etc.).

ALL studies have revealed that human subjects (both adults and infants) and even non-human primates (Newport, Hauser, Spaepen & Aslin, 2004) are indeed sensitive to non-adjacent patterns of co-occurrence, but only in specific circumstances. Gómez (2002) showed that when exposed to a simple language composed of *aXb* strings (e.g. *tep kicey rud*, presented in a sequence, separated by pauses), adult participants learned the specific dependencies between monosyllabic *a* 3and *b* elements (e.g. between *tep* and *rud*), but only when the set of intervening bisyllabic *X*s was sufficiently large. When faced with a grammaticality judgment task, subjects showed a significant preference for consistent a_iXb_i over inconsistent $*a_iXb_j$ dependencies when *X* could be instantiated by 24 different nonce words, but not 12. This suggested that the high variability of elements spanned by a dependency facilitates NAD-learning, quantified as sensitivity to the fact that the first element in a string predicts the specific form of the third element. Gómez suggested that the high variability of intervening elements rendered the adjacent transitional probabilities low (the probability of a certain token *X* following a certain token a_i, and that of a certain token b_i follow a certain *X*), and prompted the learner to ignore the adjacent patterns and look for non-adjacent regularities. Onnis, Monaghan, Christiansen & Chater (2004) showed that if the *X* item was completely invariable (set size 1) the three *a_b* dependencies were also very easily learned, and that learners not only recalled specific *aXb* strings, but could generalize their knowledge of the correct $a_i\_b_i$ dependencies to novel $a_iX'b_i$ strings with never-before-heard *X'* elements. In short, early work revealed NAD-learning to be highly constrained, and showed that **distributional cues** that differentiated between dependent elements and the intervening material facilitated learning of dependencies

Newport & Aslin (2004) showed that NAD-learning is facilitated not only by distributional cues but also by **perceptual cues**. They tested learning of dependencies in continuous strings (i.e. without separating the *aXb* sequences by pauses like Gómez, 2002), either between vowels over consonants, or consonants over vowels, or between syllables over syllables. Participants could learn segmental dependencies, but were unable to show a preference for consistent (a_iXb_i) over inconsistent ($*a_iXb_j$) strings in the test phase, when dependencies were instantiated between syllables. Results were compatible with the notion, from autosegmental phonology, that consonants and vowels are represented separately, on different phonological tiers (McCarthy, 1981), being thus grouped together and facilitating detection of patterns between (adjacent) elements of the same kind. The authors also discussed Gestalt principles of perception to account for this pattern of results: dependencies were easier learned when dependent elements were perceptually similar to each other and distinct from the intervening material. Onnis, Monaghan, Richmond and Chater (2005) showed that non-adjacent dependencies between syllables were easily learnable if phonological cues were introduced – for instance if dependent syllables began with a plosive consonant while intervening syllables began with a continuant. This strengthens the notion that NAD-learning may rely on Gestalt principles of perception that group together the dependent elements when these are similar to each other and distinct from the surrounding material.

1.3 NAD-learning as Rule-Learning or Statistical Learning

Another factor which has been proposed to contribute to the learnability of NADs is **positional cues**. Peña, Bonatti, Nespor & Mehler (2002) exposed adults to a language similar to Newport & Aslin's (2004): a continuous stream of speech presented non-adjacent regularities (a_iXb_i), such that transitional probabilities between adjacent syllables were 0.33 (a_iX, or Xb_i), but certain pairs of non-adjacent syllables exhibited transitional probabilities of 1.00 ($a_i\_b_i$). Peña and colleagues tested their subjects not on their sensitivity to specific correspondences between a_i and b_i items, but rather to their ability to recognize the a_iXb_i structures in the continuous streams, expecting that occurrence of any *a* item predicted the non-adjacent occurrence of a subsequent *b* item. Thus, participants exposed to such continuous streams for 10 minutes preferred familiar 'words' (a_iXb_i strings that had been heard as such in the input) to 'part-words' (b_ja_iX), but were not able to show that they could generalize their knowledge of the dependencies by preferring a 'rule-word' ($a_iX'b_i$, with *X'* either an a_j or a b_j , a string that had not been heard as such during familiarization but respected the NAD) over a part-word (b_ja_iX, which had been heard as such during familiarization but did not contain a dependency). Furthermore, after 30 minutes of exposure to the same type of stimuli, participants actually preferred part-words over rule-words. The authors explained these results by suggesting that participants exposed to the continuous streams were computing statistics between syllables, so that in time their preference for statistically more likely strings increased (like part-words), but that they had not learned the NADs as generalizable rules ('whenever you hear a member of the *a* set of words, predict the non-adjacent occurrence of a subsequent member of the *b* set'). By contrast, when subtle, barely perceptible (25ms) pauses were inserted at the edges of a_iXb_i strings, participants preferred both words *and* rule-words over part-words, the latter even when the familiarization was as short as 2 minutes. The authors concluded that a different learning mechanism, concerned with detecting structural rules, was operational when subtle segmentation cues were present.

Endress & Bonatti (2006) further argued for a discrepancy between statistical learning on the one hand, and rule-learning on the other: they proposed that learners were computing co-occurrences statistics between specific tokens, in both segmented and unsegmented speech. With extended exposure, this strengthened the recognition of familiar groups of syllables (like, for instance, part-words), and also allowed the detection of non-adjacent dependencies that induce preference of 'rule-words' ($a_iX'b_i$) over 'class-words' ($a_iX'b_j$, where the dependency is violated but the correlation between the structural positions and the *a/b* classes is maintained, i.e. an a element is always string-initial, and a b element is string-final). However, they proposed, in the initial stages of learning, prior to the detection of statistical regularities another mechanism helped detect what the authors call 'structural regularities' (inducing preference of class-words over part-words). Unlike the statistical learning mechanism, this mechanism relied on segmentation cues. Endress & Mehler (2009) showed that, in effect, what this rule-learning mechanism was sensitive to were positional cues: class-words were preferred to part-words when the $a_i\_b_j$ configuration was instantiated at string edges (in string-initial and –final position, a_iXYZb_j), but not when they were embedded within strings (Xa_iYb_jZ). Thus, the segmentation cues in Peña et al. (2002) and Endress & Bonatti (2006) merely served to situate the target elements in highly salient positions (string-initial and string-final), where they would be easier to encode. Other studies showed that similar types of rule-learning (of phonotactic rules, Endress & Mehler, 2010, repetition-based structures, Endress, Scholl & Mehler, 2005, etc.) were also sensitive to positional salience, and therefore easier to acquire when instantiated at string edges.

This line of research initiated by Peña and colleagues and continued by Endress and colleagues shows in very fine detail how changing the items on which subjects are tested can change the specific type of sensitivity that is being investigated. Thus, pitting words against class-words (such as Gómez, 2002) will measure knowledge about the specific one-to-one dependencies between *a* and *b* items, whereas pitting *rule*-words against class-words (the way Onnis et al., 2004 do) will measure the ability to generalize this (dependency) knowledge to novel contexts. On the other hand, pitting (rule-) words against part-words might reveal a more abstract representation of the *a* and *b* classes of items and their positions in a string relative to each other (without attention to the specific properties of their individual items). It is important to remember that even in highly controlled and simplified ALL experiments learners may be deriving a variety of representations and rules from the input and each of these representations can be evidenced in a different testing method.

If the detection of remote dependencies can involve either an abstract rule-learning process or an item-specific, statistical learning one, and if these are different processes influenced by different types of cues, then it is important to study them independently. In this dissertation I focus on the human ability to detect item-specific, one-to-one mapping between dependent elements at a distance, that is, the knowledge that, for instance, the specific syllable *ba* predicts the occurrence (over an intervening syllable or group of syllables) of the specific syllable *po* and not the occurrence of the syllable *ki*. Endress & Bonatti (2006) dub this type of learning *statistical learning* as it involves detecting relationships between specific items and not between more abstract representations (e.g. generalizing over classes of items). This type of learning seems to be uninfluenced by positional cues (e.g. marking the dependent elements as string-peripheral, therefore positionally salient, cf Endress & Bonatti, 2006), but can be guided by phonological cues marking the similarity between the dependent elements and their distinctiveness from the intervening material (Newport & Aslin, 2004; Onnis et al., 2004), or distributional cues drawing learners' attention from adjacent to non-adjacent co-occurrence probabilities (Gómez, 2002). In short, learners may detect the high co-occurrence probabilities between specific items in the input, but only in specific circumstances, where certain cues (phonological, distributional, etc.) draw the learner's attention to the target non-adjacent elements simultaneously (see Pacton & Perruchet, 2008).

1.4 The Nature of NAD-Learning

From the studies discussed so far, therefore, it seems that learners exploit a combination of distributional, phonological, and other cues to detect NADs, and do not simply compute co-occurrence statistics between any non-adjacent elements. Is NAD-learning then statistical in nature, and if not, what is the nature of the mental computations that allow learners to detect NADs? Newport & Aslin (2004) proposed that NAD-learning entails the same type of computations as argued in Aslin, Saffran & Newport (1998): learners simply tune into the transitional probabilities between elements (vowels/consonants, syllables, words, etc.) whether or not these elements are adjacent to each other. Thus, in the string *ABCDEF*, learners will be tracking co-occurrence probabilities between *A* and *B* but also *A* and *C*, *A* and *D*…, *B* and *C*, *B* and *D*, etc. However, Newport and Aslin (2004) argued, the longer the strings and the more varied the input, the more cognitively taxing it might prove to keep track of all possible co-occurrence relations. For every string with n elements, there would be n*(n-1)/2 potential dependencies, (n-1) of which would be adjacent while the rest (n-2)*(n-1)/2 would be non-adjacent; if we add to that the notion that such computations can apply at all linguistic levels – phoneme, morpheme/syllable, word or constituent, the amount of information that needs to be retained and processed in order to uncover potential regularities might become unmanageable.

Thus, Newport & Aslin (2004) proposed that while NAD-learning relies on the computation of transitional probabilities (TPs, similar to Saffran et al., 1996), calculating TPs between *all* possible remote dependencies leads to a computational overload. For this reason, they propose that the statistical learning mechanism is guided by cues like Gestalt principles of similarity: according to these principles, elements that are similar in the input are grouped and processed together even when they are linearly non-adjacent. Namely, in a sequence *ABCD*, if and *A* and *C* share certain features that are not shared by *B* and *D*, the former are grouped together as if they were adjacent. By these principles a statistical learning mechanism will calculate TPs between *A* and *C*, but not, for example, between *A* and *D*. In other words, NAD-learning may be construed as a two-part process: one in which a variety of perceptual or distributional cues can highlight elements in the input that might entertain a relationship, and a subsequent statistical learning process in which the co-occurrence probabilities of the previously highlighted elements are computed.

That NAD-learning may include an initial stage where dependent elements are selectively highlighted in the input through perceptual or distributional cues is suggested by the breakdown in NAD-learning when those cues do not guide the learner towards the target dependent elements. That the subsequent stage consists of computing co-occurrence *probabilities* is more difficult to demonstrate. Most ALL studies investigating NAD-learning use so-called 'deterministic' dependencies, in which the elements forming the *a_b* dependencies predict each other with a 100% probability, while the incorrect dependencies are formed of elements that never co-occur. From such designs, however, it is not possible to infer whether learners are actually sensitive to transitional *probabilities*, or simply to the *frequency* with which two elements co-occur. Evidence that NAD-learners are sensitive to co-occurrence *probabilities* between the remote dependent elements comes from a study by van den Bos, Christiansen & Misyak (2012). They employed a set-up similar to Gómez (2002) to compare the learning of 'deterministic' dependencies with 'probabilistic' dependencies, which differed only in the transitional probability (100% vs. 50%, respectively, with absolute frequency held constant) of the *a_b* dependencies. Probabilistic dependencies were shown to be harder to acquire than deterministic ones, requiring additional (visual or phonological) cues to facilitate learning (note however that learning was measured as discrimination of correct and incorrect dependencies *in familiar strings*, therefore not probing the ability to generalize). That learning is more robust when the transitional *probability* of *a_b* dependencies is higher (with frequency kept constant) suggests that NAD-learning indeed relies on probabilistic information, and is in this similar to other statistical learning mechanisms such as the one shown in Saffran et al. (1998) (see also Vuong, Meyer & Christiansen, 2011; Uddén, Ingvar, Hagoort & Petersson, 2012).

Not only can learners tune into the probability with which an *a* element predicts a subsequent *b* element of a dependency, but they can use this predictive power in online processing. Misyak, Christiansen & Tomblin (2010) presented participants with a spoken a_iXb_i language and simultaneously tasked them with finding and clicking the words that were being spoken in an array on a computer screen. As participants received more input, their identification of the final b_i word of the string became faster when it was accurately predicted by the a_i, but when inconsistent a_iXb_j strings were presented participants were significantly slower at identifying the b_j word. A final string-completion task also showed that participants presented with the first two words a_iX of a string could predict with a high rate of accuracy the final b_i word. Accuracy in this final task also correlated with speed of processing of complex syntactic structures such as object relative clauses (e.g. *The reporter that the senator attacked admitted the error*, see also Mysiak & Christiansen, 2012). Thus, the type of processing involved in NAD-learning may be related to the processing of complex linguistic structures, especially those that entail a syntactic relationship between distant elements (e.g. the subject of the main clause which is also understood as the object of the subordinate clause verb).

1.5 NAD-Learning in Development

Infants from around the age of 15 months are capable of detecting NADs in the same type of ALL experiments as adults. Gómez & Maye (2005) employed an artificial grammar similar to that of Gómez (2002) to test infants' sensitivity to remote dependencies at 12, 15 and 18 months respectively. Infants were exposed to an *aXb* language for 3 minutes, and subsequently tested in a Headturn Preference Procedure (cf. Kemler-Nelson et al., 1995) for their preference for test items that were consistent (a_iXb_i) or inconsistent ($*a_iXb_j$) with the exposure language. In a Headturn Preference Procedure the child is placed on a caregiver's lap and fixates on a blinking light in front of her. Once the child has fixated, the light in front is turned off, and a light on the left or right side starts blinking; when the child orients towards the respective light at an angle of at least 30°, the test trial is initiated, consisting of stimuli that are either consistent or inconsistent with the familiarization language. When the child looks away for more than 2 seconds, a new test trial begins. Across the test phase, the looking behavior (i.e. looking/listening time) is averaged over consistent vs. inconsistent trials: if a significant difference between the two types of trials emerges, then the child's behavior indicates an ability to discriminate between the two types of stimuli. This is taken to signify that learning has occurred in the familiarization phase, and that the infants discriminate the test stimuli based on the knowledge they obtained at familiarization. The group may show longer looking times towards consistent stimuli (familiarity preference) – indicating an incipient sensitivity to the relevant pattern - or, on the contrary, towards inconsistent stimuli (novelty preference, see Hunter & Ames, 1988; Houston-Price & Nakai, 2004).

Gómez & Maye found a significant familiarity preference in 15-month-olds, and a significant novelty preference in 18-month-olds, suggesting that the ability to detect NADs may have matured between these ages. Crucially, and similar to Gómez's (2002) results with adults, this preference was only found when the intervening *X* element was sufficiently variable as to highlight the invariance of the dependent *a/b* elements. Thus, infants seem to exhibit the same constraints in NAD-learning, suggesting that they might be applying the same learning strategies to the task. Twelve-month-olds, on the other hand, even with very high variability in the intervening material did not exhibit a significant difference in looking times between consistent and inconsistent trials. Gómez and Maye proposed that tracking remote dependencies puts a strain on (working/short-term) memory resources because it requires the learner to keep the *a* element in mind while processing the intervening material *X*, in order to be able to connect it to the subsequent *b*. Twelve-month-olds, unlike older infants or adults may not have sufficient working memory resources to track dependencies at a distance.

Indeed NAD-learning does seem to depend on working memory limitations: even adults exhibit limitations in the distance over which they can track dependencies (Grama, Wijnen & Kerkhoff, 2013), or the number or configuration of the dependencies they can track at one time (de Vries et al. 2012), and infants show distance limitations in tracking dependencies in natural languages as well (Santelmann & Jusczyk, 1998). Furthermore, Lany & Gómez (2008) showed that even twelve-month-olds could detect remote dependencies between classes of elements (noting that in an a_icX_i string two a_i elements combined with 8 X_i elements, while two different a_j elements combined with 8 different X_j elements), as long as they were pre-familiarized with the dependencies as adjacent (i.e. they were first presented with a_iX_i strings). Thus, one's processing resources may be taxed by the distance between non-adjacent dependent, but if this distance is rendered smaller or introduced gradually, NAD-learning will improve.

While behavioral evidence seems to suggest that infants become capable of tracking remote dependencies in the input only after their first year of life (and a few months before they exhibit sensitivity to remote dependencies in their native language, see Santelmann & Jusczyk, 1998), a different method of testing appears indicative of NAD-learning in much younger populations. Friederici, Mueller &

Oberecker (2011) investigated event-related brain potentials (ERPs) of 4-month-old German infants after exposure to dependencies in Italian:

(1) a. *La sorella sta cantando.*

The sister is singing

b. *Il fratello puo cantare.*

The brother can sing __

The infants were exposed to the dependencies for more than four times longer than in Gómez & Maye (2005) – 13.2 minutes. The experimenters tracked electrophysiological responses to incorrect (scrambled) dependencies (**sta cantare*, **puo cantando*) and compared them to responses to the correct dependencies. Importantly, the dependencies spanned monosyllabic verb stems, a distance that is also shorter than the bisyllabic words that Gómez & Maye (2005) used. Upon hearing the suffix that was not predicted by the auxiliary or the modal verb, infants exhibited a significantly different response (a late positivity identified as a P600) than they did when the predicted suffix occurred. Furthermore, this difference did not emerge in the first few minutes of exposure, but was significant in the last of the four learning blocks, indicating that infants were only tuning into the regularities after prolonged exposure. In short, infants of as little as 4 months may be able to track co-occurrence patterns between non-adjacent elements in linguistic input (see also Mueller, Friederici & Männel, 2012 for evidence that even some 3-month-old infants show learning of remote dependencies but that this ability correlates with their pitch perception). However, they may require a larger amount of input and a shorter distance between the dependent elements. As infants grow older their NAD-learning abilities may mature, allowing them to extract NADs faster and across longer intervening material.

To sum up the points made so far, humans (both adults and infants) are able to track co-occurrence patterns not only between adjacent (Saffran et al. 1996; Saffran et al., 1998) but also non-adjacent (Gómez, 2002) elements in the input. In both cases this ability seems to rely on the probability with which an element *a* predicts the subsequent occurrence of an element *b* (either immediately or over some intervening material *X*; see Saffran et al., 1998; van den Bos et al., 2012), which has prompted some authors to conclude that adjacent and non-adjacent dependency-learning engage the same basic mental computation (Newport & Aslin, 2004; Pacton & Perruchet, 2008; Vuong et al., 2011; Uddén et al., 2012).

On the other hand, tracking non-adjacent dependencies has proven to require additional cues marking the dependent elements as similar to each other but distinct in their environment (Gómez, 2002; Newport & Aslin, 2004; Onnis et al., 2005). Thus, while co-occurrence probabilities are computed between both adjacent and non-adjacent elements, it seems that statistical computations between non-adjacent elements are not performed automatically (as between adjacent elements) but need to be guided by additional cues.

This indicates that NAD-learning is distinct from adjacent dependency-learning (without its own constraints and limitations), although it also relies on probabilistic computations to identify dependencies. Like adjacent dependency-learning, NAD-learning can apply to non-linguistic input as well: humans can learn dependencies between tones (Creel, Newport & Aslin, 2004; Endress, 2010) or even non-linguistic non-musical noise stimuli (Gebhart, Newport & Aslin, 2009). However in these domains as well additional cues are required to mark the similarity between the dependent elements and their distinctiveness in the environment.[2] In other words, while the *computation* performed in adjacent and non-adjacent dependency learning may be the same (computing co-occurrence probabilities between elements), the *mechanisms* of non-/adjacent dependency learning (the succession of mental processes that engender a sensitivity to the dependencies) are crucially different. The NAD-learning mechanism is unique in that it requires specific cues that may direct the statistical computations selectively towards the target dependent elements in the input. In the following section I discuss whether this NAD-learning mechanism could be relevant to the type of remote dependencies that are frequently found in linguistic input.

1.6 Dependencies in Natural Languages

Natural languages display patterns of co-occurrence between linearly non-adjacent elements at various linguistic levels (phonological (2), morphological (3) or syntactic (4)), and these co-occurrence patterns indicate an abstract and meaningful rule of grammar:

[2] Non-human species themselves can detect NADs in a variety of modalities, both visual and auditory (Ravignani, Sonnweber, Stobbe & Fitch, 2013; Sonnweber, Ravigniani & Fitch, 2015), although they may not exploit the same cues to NAD-learning as adults (Newport et al., 2004).

(2) a. ip // ip**ler** (Turkish)

‘rope’ **Nom.sg** //**Nom.pl**

b. el // el**ler**

‘hand’ **Nom.sg** //**Nom.pl**

c. pul // pul**lar**

‘stamp’ **Nom.sg** //**Nom.pl**

d. son // son**lar**

‘end’ **Nom.sg//Nom.pl**

The examples in (2) illustrate the phenomenon of **vowel harmony** in Turkish. The nominal inflection for plural is *-ler* when the vowel preceding it is a front vowel (*i* or *e*), but changes to *-lar* when the preceding vowel is a back vowel (*u* or *o*). In fact, these surface co-occurrence patterns reflect a more abstract and widely encountered phonological rule in Turkish, stating that the (last) vowel in the noun transfers its feature(s) (in this case [+/-back]) to the vowel of the suffix. Such a rule could be inferred by observing the high probability of co-occurrence of vowels with the same phonological features ([+/- back]), irrespective of the consonants intervening between the respective vowels.

(3) a. k**a**t**a**b (Hebrew)

write.**3rd.masc.sg.perfect** ‘he wrote’

b. **yi**kt**o**b

write.**3rd.masc.sg.imperfect** ‘he writes/will write’

c. k**o**t**e**b

write.**masc.sing.active participle** ‘writer’

Example (3) illustrates consonantal roots in Semitic languages: the lexical root is a template formed of consonants alone; the vowels that fill this template are meant to inflect the root for person, gender, number, tense, etc. Detecting the dependency between specific consonants over vowels (i.e. *k-t-b*) by observing the high co-occurrence rate (in different contexts) between these consonants, allows the learner to extract the lexical root; detecting the correspondences between specific vowels (over different consonantal roots) allows the learner to acquire the inflectional rules.

(4) a. Noi toţi greşim câteodată. (Romanian)

We all err.1st pl sometimes

a'. Voi toţi greşiţi câteodatã.

You all err.2nd pl sometimes

a''. * Voi toţi greşim câteodată.

b. Ik heb vandaag de dokter gebeld. (Dutch)

I have today the doctor PART.call.PART

'I have called the doctor today'

c. una ragazza bella // un ragazzo bello (Italian)

a.fem girl.fem beautiful.fem // a.masc boy.masc beautiful.masc

Dependencies in natural languages are not only formed of sub/segmental units. Full morphemes can also enter dependencies, and when they do, as in (4), they mark important morpho-syntactic relationships. As shown in (4a, a'), the form of the subject pronoun in Romanian predicts the form of the suffix on the verb, such that there is a one-to-one correspondence between the two. One cannot pair a subject pronoun with a different suffix than the one it predicts without making the sentence ungrammatical (4a''). This co-occurrence pattern is the manifestation of a formal relationship that obtains between a subject and the main verb of its clause: agreement, namely the fact that the (person, number) features on the verb must match the (person, number) features on the subject. Agreement can also manifest itself at the level of the nominal phrase: in (4c) the article and the adjective agree in gender with the noun, a relationship that is reflected in the co-occurrence pattern of the gender suffixes on the three words. Note that all dependencies exemplified in (4) obtain between functional morphemes (auxiliaries, determiners, inflection, etc.), and that these morphemes may be either free (auxiliaries, determiners) or bound (gender suffix, participle prefix, etc.). Note, also that the intervening material between the dependent elements may contain one or more morphemes, and can be highly variable, suggesting that the dependencies may be optimally tracked by a learning mechanism that abstracts away from the intervening material.

Other important dependencies in natural languages can reflect verbal aspectual paradigms: in (4b) the auxiliary *have* in Dutch combined with the participle circumfix mark the perfective aspect of the action denoted by the verb (i.e. the fact that it is completed prior to a reference point in time, in this case the reference point being the moment when the sentence is uttered). *Have* can also occur on its own, without the participle, when it is a lexical verb (expressing possession) instead of an auxiliary; and the *ge-t/d* circumfix can also occur on its own in adjectives like *gevlekt* (‘spotted’). Yet the interpretation of the verb as perfective only obtains when the two occur together.

This dissertation focuses exclusively on dependencies similar to the ones in (4). A naïve learner exposed to examples as in (4) may have no pre-existing knowledge that, for instance, the pronoun subject in (4a) expresses the features 1st person plural, nor that the suffix on the verb must carry the same features. However, the learner will be able to notice, with consistent input, that the pronoun predicts (with a high probability) the occurrence of the correct suffix (and vice-versa), and not of some other suffix like in (4a''). As we have seen in section 1.2, learners are capable of tracking these relationships, and prefer dependencies where the second element is predicted by the first to dependencies where it is not. With exposure to the whole paradigm (examples (4a), (4a'), etc.) the learner might infer that the morphological forms of the items in those particular positions (i.e. subject and verb-suffix) are always correlated, and that, therefore, there must be a syntactic relationship between the two items. Thus, by observing surface properties of the input such as co-occurrence patterns between specific items, one could arguably infer more abstract morpho-syntactic rules of natural languages. This dissertation explores exactly this possibility: that a learning mechanism capable of tracking co-occurrence patterns between non-adjacent elements might aid the ‘naïve learner’ (the infant acquiring their native language) in acquiring morpho-syntactic dependencies like the ones in (4).

1.7 Learning Morpho-Syntactic Dependencies in Natural Languages

For infants learning their native language, sensitivity to morpho-syntactic dependencies such as the ones in (4) is attested as early as 18 months. Santelmann & Jusczyk (1998) tested English-learning 15-month-olds and 18-month-olds on their ability to detect the dependency *is_ing* representing the progressive aspectual paradigm (marking the fact that the main verb reflects an ongoing action at the time of utterance). In a Headturn Preference Procedure (with no prior familiarization), infants were exposed to trials consisting of passages with either correct (3rd person singular auxiliary *is* and the suffix *–ing*, cf. 5a) or incorrect dependencies (the modal verb *can* replaced the auxiliary, cf. 5b):

(5) a. Out in the desert the archaeologist is digging for treasures. One young worker is helping her. An old worker is bringing them cold water to drink. That scientist is looking at some old pots. In the tent, someone is putting a necklace in a new case. Everyone is trying to be careful.

b. Out in the desert the archaeologist can digging for treasures. One young worker can helping her. An old worker can bringing them cold water to drink. That scientist can looking at some old pots. In the tent, someone can putting a necklace in a new case. Everyone can trying to be careful.

Across trials, 18-month-olds (but not 15-month-olds) oriented longer to passages exhibiting the correct dependencies, like (5a), than to passages where the modal verb formed an incorrect dependency with the *–ing* suffix. Subsequent studies showed that in other languages as well (German, Dutch, French) children around 18-24 months showed sensitivity to morpho-syntactic dependencies (Wilsenach and Wijnen 2004; Höhle et al. 2006; van Heugten and Johnson 2010; van Heugten and Shi 2010). This is relevant if we recall that, within the same experimental paradigm (the Headturn Preference Procedure) Gómez & Maye (2005) showed that 15-month-old infants were detecting NADs in unfamiliar input from minimal (3-minute) exposure.[3] In other words, behavioral evidence suggests that infants' ability to learn remote dependencies matures just as they start to exhibit a sensitivity to morpho-syntactic dependencies in their native language, although more evidence is needed to establish a causal link between these two observations.

It has been proposed that the NAD-learning mechanism may promote the acquisition of morpho-syntactic dependencies by facilitating the detection of one-to-one mappings between dependent elements in the input. Although it is difficult to obtain direct evidence that infants acquiring their native language are using a specific type of mental computation to make sense of the input, several indirect methods provide proof of concept that NAD-learning may be linked to morpho-syntactic dependency-learning. For instance, the working memory limitations that seem to constrain NAD-learning (Lany & Gómez, 2008; Grama et al., 2013) also affect the acquisition of morpho-syntactic dependencies. Thus, Santelmann &

[3] Furthermore, Lidz, Omaki & Orita (2012) showed that even 15-month-olds could show sensitivity to the dependency in Santelmann & Jusczyk's sentences, provided that an additional training phase presented them with repeated instantiations of the dependency.

Jusczyk (1998) also showed that 18-month-olds could track the respective dependencies across larger distances, when a bisyllabic adverb intervened between *is* and *–ing*, in addition to the monosyllabic verb stem, (e.g. 'Out in the desert the archaeologist is *keenly* digging for treasures.'). However, when the distance between the dependent elements was increased to 5 syllables (e.g. 'Out in the desert the archaeologist is *quite intently* digging for treasures.'), infants no longer discriminated the correct dependencies, suggesting that they could not track non-adjacent patterns when the dependent items were too far apart. Santelmann & Jusczyk accounted for this finding by claiming that 18-month-olds had limited working memory capacity, namely that they processed the input by segmenting it into small chunks; when more than 3 syllables intervened between dependent elements, the two members of a dependency could no longer be processed in the same chunk, and thus, the relationship between them was lost.

Höhle, Schmitz, Santelmann & Weissenborn (2006) showed that distance in number of syllables is not the only factor that influences the processing of remote dependencies: 19-month-old German children were not able to detect dependencies between the *have*-auxiliary and the participle in German with an intervening bisyllabic adverb (*Das kleine unzufriedene Kind hat/*kann **morgens** geheult* 'The little unhappy child has/*can cried in the morning') but were perfectly able to do so with an intervening bisyllabic noun phrase (*Das kleine phantasievolle Kind hat/*kann **den Ball** geholt* 'The little imaginative child has/*can the ball fetched'). They argued that the determiner of the noun phrase may represent a reliable marker helping the children label the intervening material as a noun phrase; at the same time, the adverbs were not marked in any reliable, recognizable way (as were the English adverbs in Santelmann & Jusczyk's study, for instance, which predominantly exhibited the adverbial suffix *-ly*). Thus, the authors claimed, the working memory limitations that constrain the processing of morpho-syntactic dependencies may not so much be linked to the *length*, in syllables, of the intervening material, but rather to the speed and efficiency with which this material can be processed. Intervening material that is slow or difficult to process will disrupt the processing of the remote dependency.

Höhle et al.'s (2006) study also draws our attention to the fact that even if we assume morpho-syntactic dependencies are detected based on a purely statistical learning mechanism, this process might rely on the acquisition of other structural properties of the language, such as morphological class, functional elements and their combinatorial properties, etc. Thus, the acquisition of remote morpho-syntactic dependencies may rely both on the growth and sophistication of the NAD-learning mechanism, as well as on the acquisition of (lower-level) aspects of morpho-syntax, such word-class and adjacent dependencies.

Another indirect way to establish a link between NAD-learning and the acquisition of morpho-syntactic dependencies is to correlate the age of acquisition of certain dependencies with their distributional properties in corpora of child-directed speech, in order to substantiate the claim that a distributional learning mechanism is responsible for their acquisition. Van Heugten & Johnson (2010) showed that Dutch 24-month-olds' performance at tracking two distinct determiner-suffix dependencies within the nominal domain (*het_-je* and *de_-en*) patterned with the mean frequency, co-occurrence probability and relative proximity (average length spanned by the NAD) of these dependencies in a corpus of child-directed speech. The children were sensitive to the dependency between the neuter determiner *het* and the diminutive suffix *–je*, but did not track the dependency between the determiner *de* and the plural suffix *–en*, the former being more frequent, having higher transitional probability and generally spanning shorter distances than the latter.

Tincoff, Santelmann & Jusczyk (2000) extended the findings of Santelmann & Jusczyk (1998) by investigating 18-month-olds sensitivity to various other forms of the *be_ing* aspectual paradigm, besides the 3rd person singular *is_-ing*: i.e. *was/are/were_-ing*. They discovered a sensitivity to the dependency between the past tense 1st/3rd person singular *was* and the *–ing* suffix, which infants preferred over the incorrect **could_-ing*, but no sensitivity to the *are_-ing* or *were_-ing* dependencies (which were not significantly preferred over the incorrect **will_-ing* and **would_-ing*, respectively). However, the authors could not match the pattern of 18-month-olds' sensitivity to these dependencies to their frequency and co-occurrence probability in a corpus of child-directed speech; for instance, *was_-ing* was much less frequent than *are_-ing*, but infants were more sensitive to the former than to the latter. One tentative explanation for these results could be that infants collapse *was_-ing* and *is_-ing* together, especially due to the phonological similarity between the two; research shows that phonological detail on functional morphemes is hard to acquire due to their lack of perceptual salience (cf. Chapter 2 for discussion and references); if *was_-ing* is initially treated as *is_-ing* and then only subsequently differentiated phonologically, then *was_-ing* could piggy-back on the acquisition of *is_-ing*, explaining why it is known to 18-month-olds despite its highly infrequent nature. On the other hand, *are_-ing* and *were_-ing*, even cumulatively, are quite infrequent. By this account, we can restore the notion that the acquisition of *be_-ing* is highly input-dependent.

Other studies have brought indirect evidence of a link between NAD-learning in artificial grammar paradigms and acquisition of morpho-syntactic dependencies in natural languages, by showing that certain (language-impaired) populations who have deficits in one also have deficits in another. Thus, Wilsenach & Wijnen (2004) showed that, around the age of 20 months, typically developing Dutch children are sensitive to the have+participle (perfective aspect) dependency in Dutch (*Het heeft vandaag geregend* 'It has today PART.rained'), whereas age-matched children at genetic risk for dyslexia (and associated language learning difficulties) fail to exhibit sensitivity to this dependency.

Kerkhoff, de Bree, de Klerk & Wijnen (2013) replicated Gómez & Maye's (2005) NAD-learning study with 18-month-old Dutch infants, typically developing and at familial risk for dyslexia. As in Gómez & Maye's study, typically developing infants exhibited a novelty preference at test – significantly shorter looking times towards test items consistent than to those inconsistent with the dependencies they had heard during familiarization. The group at risk for dyslexia, however, did not display a significant preference in either direction, suggesting that the NAD-learning mechanism in the at-risk population was somehow impaired or delayed. This result is in line with other studies indicating a more general implicit learning deficit in children at risk for dyslexia (Pavlidou, Kelly & Williams; 2010). Lack or delay of sensitivity to morpho-syntactic dependencies at 20 months could, therefore, be explained by problems with NAD-learning at 18 months, although it is the task of future research to substantiate the claim that these abilities are in fact correlated.

To briefly summarize, the acquisition of morpho-syntactic dependencies in natural languages shows some connections to the development of a (statistical) learning mechanism that detects dependencies between non-adjacent elements in the input. Behavioral evidence suggests infants master NAD-learning around the same age that they show familiarity with morpho-syntactic dependencies. NAD-learning and processing morpho-syntactic-dependencies exhibit similar working-memory limitations whereby dependencies are harder to detect when the intervening material is longer or more difficult to process. The order in which morpho-syntactic dependencies are acquired could be linked to the distributional properties (frequency, transitional probability) of those dependencies in the input, the type of distributional properties exploited by a mechanism relying on statistical computations (like NAD-learning). Finally, in populations where statistical learning mechanisms like NAD-learning are impaired, the acquisition of morpho-syntactic dependencies is also delayed.

1.8 Purpose of the Current Study

As seen in the previous section, evidence from a variety of lines of investigation converges to show similarities between NAD-learning and the acquisition of morpho-syntactic dependencies. A different approach to this question has been less exploited in ALL studies: namely, the extent to which properties of morpho-syntactic dependencies in natural languages could, in principle, facilitate NAD-learning (or, vice-versa, the extent to which constraints placed on NAD-learning as identified in carefully controlled ALL studies are consistent with cues available across natural languages). Gómez (2002), for instance, pointed out that the elements that usually form morpho-syntactic dependencies are part of the class of functional morphemes in a language: pronouns, determiners, auxiliaries, verbal and nominal inflection, etc. These elements are quite frequent in the input and they are fairly invariant, with a largely grammatical function. Elements that are spanned by such dependencies, verbs, nouns, adverbs, adjectives, etc., are part of the class of lexical words, not as frequent and much more variable, part of a large and ever-expanding set. Gómez (2002) showed that NAD-learning was optimal with dependencies between frequent and invariant elements, spanning highly variable elements. Thus the distributional properties of natural languages seem to be promote the detection (by a statistical learning mechanism) of dependencies between functional morphemes.

In the current study I adopt an approach similar to Gómez's (2002) towards answering the research question: could NAD-learning, in principle, subserve the acquisition of morpho-syntactic dependencies in natural languages? If infants rely on NAD-learning to detect dependencies between functional elements in natural languages, then the properties of these elements, as well as the timeline of their individual acquisition, should facilitate NAD-learning. In other words the constraints on the NAD-learning mechanism should be satisfied by the properties of the (natural language) input[4]. This question can be investigated in an ALL paradigm where, by carefully controlling the properties of the input, learning can be compared between conditions emulating the properties of natural languages and those that do not. In the chapters that follow, therefore, I employ an ALL paradigm to investigate how NAD-learning could be guided by cues available in the input or intake of infants acquiring morpho-syntactic dependencies in a natural

In the first two chapters I use adult learners as a model for infant learning, and in the final chapter I investigate whether infants show similar learning abilities to adults. ALL studies with adults have often been used as a model for language acquisition, since learners are faced with unfamiliar input and learn incidentally, without any explicit instruction (and often without explicit attempts to identify regularities). Adult studies have a number of practical advantages: they are easier

[4] Note that this is not a claim as to whether language was shaped by the demands of the NAD-learning mechanism, or vice-versa. Rather, if NAD-learning is to be successfully applied to linguistic input, then the constraints on this learning mechanism should be met by that linguistic input. Newport & Aslin (2004) suggest that the specific properties of remote dependencies in natural languages may have arisen from the constraints on NAD-learning, but this is orthogonal to the research question of this dissertation.

to set up, faster to run and provide more direct measures of learning (i.e. more easily interpretable results such as the proportion of preference for grammar-consistent over inconsistent stimuli at test). These studies provide proof of existence of a certain learning mechanism and its properties, and make predictions for the types of effects that should be observed with infants. I follow this pattern in investigating features of NAD-learning in adults first, and then attempting to replicate some of these findings with infants as well.

Therefore, in Chapter 2 I build on the observation that functional categories in natural languages are not just marked by distributional cues that distinguish them from lexical categories, but also by a constellation of perceptual/prosodic cues. I therefore set out to investigate how the perceptual/prosodic cues that mark functional elements (in at least some languages) may guide the NAD-learning mechanism.

Chapter 3 relies on an observation drawn from behavioral studies of the acquisition of functional morphemes and morpho-syntactic dependencies: whereas infants identify many of the functors in their language by the end of their first year, remote dependencies between these functors are only detected around the age of 18 months. This timeline suggests that individual functional elements are encoded in memory before relationships between them are detected in the input. In this chapter I investigate whether prior familiarity with the individual morphemes facilitates the detection of dependencies between them.

Finally, in Chapter 4 I turn my attention to infants around the age of 18 months. I set out to investigate whether they show the same patterns of learning as the adults in Chapter 2, being able to detect dependencies and generalizing them to novel contexts, as well as showing robust learning of dependencies between prosodically non-salient elements over more salient intervening material.

The purpose of the current work is twofold. Firstly, I propose to offer evidence that there are cues in natural languages, and in the order of acquisition of various aspects of grammar, that are uniquely tuned to the constraints of the NAD-learning mechanism, such that young infants can exploit this mechanism effectively to detect relevant regularities in their native language. Secondly, in investigating the effect of certain cues on NAD-learning I propose to extend our understanding of the workings of a mechanism that has puzzled psycholinguistics for a long time.

Chapter 2

The Role of Prosodic Cues in Non-Adjacent Dependency-Learning[5]

In the previous chapter I discussed the human ability to detect co-occurrence patterns between non-adjacent elements in strings of unfamiliar speech (NAD-learning), and I hypothesized that this ability might be recruited for the acquisition of various remote dependencies in natural languages. I focused on a particular type of natural language dependencies, namely morpho-syntactic dependencies (see (1) - (3) below), and argued that the properties of these dependencies in natural languages might satisfy the constraints on the NAD-learning mechanism. Thus, Gómez (2002) showed that dependencies between words *a* and *b*, occurring non-adjacently in *a*X*b* strings, are only learned successfully when the dependent elements are frequent and stable (invariant), whereas the intervening material (the Xs) is variable. In natural languages, morpho-syntactic dependencies such as the ones in (1), (2) and (3) are instantiated between functional morphemes (auxiliaries, determiners, inflectional suffixes/prefixes, etc.), which are highly frequent, closed-class elements, and span lexical morphemes (nouns, verbs, etc.), which are highly variable, open-class elements.

(1) a. Noi toţi greşim câteodată.
We all err.1st pl sometimes
b. Voi toţi greşiţi câteodată.
You all err.2nd pl sometimes

(2) Ik heb vandaag de dokter gebeld.
I have today the doctor PART.call.PART

(3) una bella ragazza / un bel ragazzo
a.fem beautiful.fem girl.fem / a.masc beautiful.masc boy.masc

---

[5] An earlier version of this chapter was published in *Journal of Psycholinguistic Research* in February 2016 under the title 'Gleaning Structure from Sound: The Role of Prosodic Contrast in Learning Non-Adjacent Dependencies'.

Thus the constraints that limit the scope of the NAD-learning mechanism could be well-matched with the constraints on the kind of dependencies that natural languages can exhibit (Newport & Aslin, 2004). In this chapter I further explore the hypothesis that the properties of morpho-syntactic dependencies in natural languages may be optimal in supporting NAD-learning. I capitalize on the observation that the functional morphemes engaged in morpho-syntactic dependencies are marked not only by specific distributional properties (high frequency, low variability) but also by a special constellation of perceptual properties. In section 2.1 I review evidence for the perceptual cues that mark functional morphemes as distinct from lexical ones. Section 2.2 offers a timeline for the acquisition of functional morphemes, and shows that this timeline is strongly influenced by the distributional and perceptual properties of these morphemes. In section 2.3 I discuss the effect that perceptual cues to functional morphemes have on the detection of dependencies between them, and in section 2.4 I detail the rationale for the current study – assessing the role of perceptual cues on NAD-learning. Experiments 1 and 2 (sections 2.5 and 2.6) investigate this question, while section 2.7 discusses the findings of these experiments and their implications.

2.1 Perceptual Cues to Functional Morphemes

Functional words in natural languages have been suggested to differ markedly from lexical ones, not just in terms of distributional properties (such as frequency and variability), but also at the level of phonetic/phonological properties (Morgan, Shi & Allopenna, 1996): (i) functors are minimal, in that they comprise as few syllables or moras as possible; (ii) their syllable structure is also minimal, in terms of onset, coda, vowel duration and complexity, consonantal clusters, etc. (iii) as (usually) unstressed morphemes, functors are shorter, with lower amplitude, and have centralized, reduced vowels. Shi, Morgan & Allopenna (1998) analyzed child-directed speech in English, Turkish and Mandarin Chinese, and showed that functors are distinguishable from lexical items along the following parameters: frequency, utterance position, number of syllables, presence of complex syllabic nucleus (diphthongs), presence of syllabic coda, syllable duration, and relative amplitude. The authors emphasize that none of the individual cues (whether distributional, morpho-phonological or acoustic) by themselves are sufficient to reliably identify functors, but that the constellation in its entirety affords around 80-90% accuracy in categorization. The study also underlines the language-specificity of cue-sets in this respect: tonal languages like Mandarin include marked tone as a cue, whereas other languages may present phonotactic cues, such as English, where word-initial inter-dental fricatives are only voiced in function words (e.g. *the*).

Monaghan, Christiansen & Chater (2007) took four different languages (English, Dutch, French and Japanese) and showed that an even larger constellation of cues (length, syllabic complexity, manner and place of articulation of consonants within the word as well as in word-initial or onset position, vowel-to-consonant ration, vowel reduction, vowel position) significantly distinguished between functional and lexical words. They showed that the four languages exhibited only partial overlap in the cues that were significant predictors of the lexical-functional distinction, thus suggesting that each language has its specific subset of phonological cues to the class distinction, and that these cues have to be learned from input. The study also showed that combining phonological cues with distributional cues to the lexical/functional distinction yields a very high predictive power: around 95% of the items in a corpus (for either one of the four languages) can be correctly classified on the basis of all the cues.

If different acoustic cues mark functional and lexical words as different in speech, it makes sense to consider the possibility that this constellation of acoustic cues reflects differences in the prosodic status of function and lexical words. Selkirk (1996) reviews cross-linguistic evidence that function words do, indeed, have a different prosodic status than lexical words. Previous work has often remarked that function words (in a variety of languages like English, Dutch, German, Portuguese, Serbo-Croatian, etc.) can be prosodically weak, behaving as 'prosodic clitics', namely units which must be attached to a prosodically stronger host (Selkirk 1986; Berendsen, 1986; Nespor & Vogel 1986; Booij, 1996; Hall, 1999; Vigário 1999; Zec, 2002).

Less research has looked into the perceptual or prosodic properties of bound functional morphemes, such as the Romanian verbal person/number suffix in (1), the Dutch participle circumfix in (2), or the Italian gender adjectival/nominal inflection in (3). Cross-linguistically, the properties of these morphemes may vary a lot: inflectional morphemes can be sub-syllabic (e.g. plural *-s* or past tense *–d* in English) or syllabic (e.g. the verbal inflection in Romanian in (1)). In languages like Dutch or English, inflectional morphemes (affixes) are prosodically weak; in Dutch, for instance, prosodic clitics (free function words) have been claimed to be prosodically indistinguishable from affixes (bound functional morphemes, cf. Booij, 1996). In a language like Romanian, on the other hand, some inflectional morphemes can receive lexical stress, such as the inflectional suffix in (1) (*greş'im, greş'iţi*). Because it would be difficult to address this variation in its entirety, in this chapter I restrict my interest to languages where functional morphemes both free and bound are generally prosodically weak. These languages are an ideal testing ground for investigating whether lack of prosodic (or perceptual) salience could inhibit, or, on the contrary, support the detection of NADs between non-salient units.

2.2 Functors in Early Acquisition

The perceptual distinction marking the functional/lexical categorization does not go unnoticed by language-learning children: 11-month-old infants are sensitive to the properties of functional words in their language, and distinguish texts containing real functors from texts with nonce, phonologically atypical ones (Shafer, Shucard, Shucard & Gerken, 1998). In fact, even newborns can discriminate lists of functional from lexical words based on perceptual cues alone (i.e. controlling for frequency and word-length, cf. Shi, Werker & Morgan, 1999). From the moment of birth, therefore, children are sensitive to, and can capitalize on subtle perceptual cues to the open/closed class distinction.

However, saying that the perceptual properties of functors distinguish them from lexical words is not tantamount to saying that these acoustic properties facilitate early acquisition of functors. Data from production seem to suggest that the functional category is poorly represented in early speech, until about the age of 2-3 years (Bassano, Eme & Champaud, 2005; Brown & Fraser, 1963). One possible explanation is that, in fact, the prosodically weak status of functors makes them less salient (although more easily distinguishable from lexical words) in natural speech. Six month-old English and Chinese infants prefer attending to lexical rather than functional words in English, when presented with alternating lists in a (central) Visual Fixation procedure (Shi & Werker, 2001; Shi & Werker, 2003). Furthermore, in a visual Habituation-Dishabituation procedure, English infants habituated with (English) functors looked longer at lexical words in the test phase, while infants habituated with lexical words showed no recovery of attention when hearing functional words in the test phase (Shi & Werker, 2001). This suggests that 6-month-olds cannot direct their attention to functional words, even though they are able to discriminate them from lexical words, possibly because the function words are not salient enough.

Moreover, although infants have been shown to distinguish functors in their native language from 6 months onwards, this sensitivity seems closely linked to both the perceptual and the distributional properties of the functors under scrutiny. From the very early age of 6-7 months, infants are able to recognize familiar functors in speech, as evidenced by their ability to discriminate phrases/texts containing functors they have been trained on from texts containing other functors they have not heard in the familiarization (Shi, Marquis & Gauthier, 2006 – French; Höhle & Weissenborn, 2003 – German). However, this discrimination performance breaks down when the contrasting (familiar vs. novel) functors are phonologically similar (*la* and *ta*, Shi, Marquis & Gauthier, 2006), or when the target functor is infrequent in the child's input, showing that 6-month-olds' knowledge of functors is input-dependent, and their representation of even familiar ones is *phonologically underspecified*. Note that Shi et al. show infants are able to distinguish *la* and *ta* in isolation, but not within noun phrases: the lack of perceptual salience of functors may be more problematic in the context of fluent speech.[6]

By 8 months, infants recognize high-frequency (but not low-frequency) determiners in noun phrases, and use this knowledge to segment and recognize the nouns they precede (Shi, Cutler, Werker & Cruickshank, 2006; Shi & Lepage, 2008; Hallé, Durand & de Boysson-Bardies, 2008). In French, where functors are more acoustically salient, infants at 8 months already show full phonological encoding (phonologically fully-specified memory representations) of the determiners (e.g. *des*), by differentiating them from phonologically similar nonce functors (*kes*, cf. Shi, Cutler, Werker & Cruickshank, 2006). However, even later, by the age of 11 months, discrimination of French determiners is still affected by the length and familiarity of the subsequent noun, suggesting that it is more difficult to segment or identify the determiner when processing resources are claimed by a long or unfamiliar subsequent noun (cf. Hallé, Durand & de Boysson-Bardies, 2008). In English, where functors are less salient, children only fully phonologically specify determiners (as indicated by their ability to discriminate *the* from the minimally different nonce word *kuh*) by 11 months, substantiating the intuition that salience determines full phonological encoding (cf. Shi & Lepage, 2008, also Stromqvist, Ragnarsdóttir & Richthoff, 2001).

In short, between 6 and 8 months, infants become familiar with the most frequent functional items in their native language: this process continues throughout the first year of life and differs from language to language as a function of the perceptual properties of the functors in question. Across languages, infants seem to traverse a stage of incomplete phonological encoding of the functors, and, also across languages, the order of acquisition of different functors depends on their frequency as well as phonological salience. Performance on the recognition of functors also seems to vary with the complexity of the context (cf. Hallé et al., 2008).

---

[6] Furthermore, given that these studies familiarize infants with the functor tokens in isolation, it is difficult to see how much they speak to the ability of 6-7-month-olds to actually detect, segment and store functors based on their day-to-day input.

Phonological encoding of functors seems to mature around the age of 11 months. At this age, English infants can exploit the phonemic detail in unstressed syllables of words with iambic stress (weak-strong, to discriminate minimal pairs, see Johnson, 2005). German infants become able to discriminate iambic-stress words (*Vulkan*) from iambic-stress determiner+noun phrases (*der Kahn*, see Höhle & Weissenborn, 2000). By 11 months, infants also become sensitive to verbal inflections (e.g. the French infinitive morpheme *-er*), and display the ability to learn a new (nonce) inflection (*-ou*) in an experimental setting, based solely on observing its co-occurrence with a variety of verbal stems (Marquis & Shi, 2012). Infants are thus familiar with grammatical morphemes as separate from the lexical roots they combine with, and the order of acquisition for these morphemes seems to be determined in part by their frequency and combinatorial power.

From 14-16 months onward, children begin to exhibit knowledge of the combinatorial properties and the syntactic position of functors (Shady, 1996; Höhle, Weissenborn, Kiefer, Schultz & Smitz, 2004; Kedar, Casasola & Lust, 2006; Shi & Melançon, 2010; Nazzi, Barriére, Goyet, Kresh & Legendre, 2011). This knowledge emerges gradually, and also seems to be dependent on the distributional cues available in the input: in French or German, for instance, where nouns are more consistently preceded by determiners than verbs are by pronouns, 14-month-olds show knowledge of the combinatorial properties of determiners (*le/un* + Noun), but not of the combinatorial properties of pronouns (*il/*je + Verb; Höhle et al., 2004; Shi & Melançon, 2010). At 15 months, English children recognize the verbal aspectual suffx *–ing* and segment it from the root it attaches to (Mintz, 2013), even while they are not yet able to detect its co-occurrence with the auxiliary verb *be* in sentences like *He <u>is</u> bak<u>ing</u>* (Santelmann & Jusczyk, 1998).

Taken together, this line of research shows the great influence that perceptual and distributional properties of functors have on their acquisition. Distributional properties facilitate the acquisition of functors as a separate class. Phonological reduction renders functional items at the same time reliably recognizable and difficult to fully encode; furthermore, perceptual properties seem to render functors less salient to young learners than their lexical counterparts. But what implications does this have for the acquisition of grammatical dependencies between functional morphemes?

2.3 The Role of Perceptual Cues in NAD-Learning

Several studies have investigated the acquisition of morpho-syntactic dependencies and have shown that infants become sensitive to such dependencies in their native language around the age of 18-24 months (Santelmann & Jusczyk, 1998; Tincoff, Santelmann & Jusczyk, 2000; Soderstrom, 2002; Wilsenach & Wijnen, 2004; Höhle et al., 2006; van Heugten & Shi, 2010; van Heugten & Johnson, 2010). A few of these studies have tried to relate the acquisition of these dependencies to their distributional properties (frequency of the dependency, transitional probability, etc.; Tincoff, Santelmann & Jusczyk, 2000; van Heugten & Johnson, 2010). However, no study so far has investigated the effect of perceptual cues on the acquisition of morpho-syntactic dependencies.

Furthermore, few studies have looked into the potential effect of perceptual cues on the acquisition of rules or patterns in general. Several studies show that artificial grammars with closed-class nonce words which are shorter than their open-class counterparts facilitate learning the structure of the grammar (Green, 1979, Morgan, Meier & Newport, 1987 and Valian & Coulson, 1988). Cutler (1993) cites a study that specifically shows phonological reduction to be a crucial facilitatory cue in Artificial Grammar Learning (AGL): subjects performed significantly better on learning the rules of an artificial grammar (with monosyllabic words) when the frequent functor-like elements in the language were weak syllables, in contrast with the open-class elements which consisted of strong syllables; subjects performed poorer either (i) when strength-assignment was random, (ii) when all syllables were strong, or (iii) when 'functors'(frequent closed-class items) were assigned strong syllables, and 'lexical items' (open-class) weak ones. Thus, the association between specific distributional and prosodic properties seems to be important in grammar-learning tasks.

Studies investigating NAD-learning with non-linguistic stimuli have also shown that perceptual cues can affect NAD-learning in crucial ways. Creel, Newport & Aslin (2004) investigated how participants learned adjacent or non-adjacent dependencies between musical tones: learners only acquired non-adjacent regularities when perceptual cues such as pitch range and timbre marked the non-adjacent units as distinct from the other units in between. In other words, dependencies had to be singled out by perceptual cues that rendered dependent elements 1) similar to each other, and 2) different from the surrounding material. Gebhart, Newport & Aslin (2009) confirmed that NAD-learning between noise stimuli showed the same pattern. Along with Newport & Aslin (2004), these studies put forth the proposal that NAD-learning needs to be guided by *Gestalt principles of similarity*: elements that are perceptually similar to each other and different from the context are somehow perceptually 'grouped together', despite the distance and the intervening units in the linear string of input. This similarity could be expressed on different levels: acoustic (Creel, Newport & Aslin, 2004; Gebhart, Newport & Aslin, 2009), phonological (vowels across consonants or vice-versa, Newport & Aslin, 2004; syllables with a plosive across syllables with a continuant, Onnis et al., 2005), etc.

Newport and Aslin argued that by using Gestalt principles of perception a highly variable input can be organized into different perceptual 'tiers', and elements with distinct perceptual properties are processed on separate 'tiers'. Thus, dependencies between non-adjacent elements in the linear input become adjacent dependencies on a separate 'tier' and the computation of co-occurrence regularities between them becomes easier. A statistical learning mechanism that computes co-occurrence statistics between adjacent elements could operate not only in the linear input but also at the level of each and every one of the 'tiers', simultaneously detecting adjacent *and* non-adjacent dependencies.

Such an account could even be extended to explain the Gómez (2002) data: elements with similar distributional properties (high frequency and low variability) are easier to 'connect' across units that have very different distributional properties (high variability). Onnis et al. (2004) showed that the reverse situation was true as well: when the intervening X did not vary at all (there was only one X), whereas the dependent elements were slightly more variable (there were 3 *a_b* dependencies), performance was also significantly above chance. Thus, learning did not depend on the specific properties of the dependent elements *per se*: the crucial criterion seems to have been simply the contrast between the properties of the dependent items and those of the intervening items.

Newport & Aslin's hypothesis could easily be applied to the problem under scrutiny here: how do the perceptual properties of functional elements in natural languages affect the ability to detect morpho-syntactic patterns?

2.4 Purpose of the Current Study

The Gestalt principles of perception hypothesis, put forth by Newport & Aslin (2004), predicts that the specific perceptual cues that mark functional words/morphemes in natural languages will facilitate the detection of co-occurrence patterns between 'functional'-sounding elements (over 'lexical'-sounding ones) in a controlled learning environment such as an AGL paradigm. In this study I employ just such a paradigm to investigate the role of perceptual distinctiveness in the detection of non-adjacent dependencies.

The hypothesis under scrutiny – **Hypothesis 1** – states that dependencies are learned based on Gestalt principles of perception: elements that are perceptually distinct in their environment, but similar to each other are grouped together (or represented together on a separate level, see Newport & Aslin, 2004), and therefore patterns between them are more easily detected. Hypothesis 1 predicts that the detection of dependencies between perceptually 'reduced' (functional) morphemes, spanning 'lexical'-sounding material, should be facilitated by the perceptual distinctiveness of functors. According to this hypothesis, then, learning dependencies in natural languages should also be enabled by the specific perceptual distinction between the functional/lexical class, which allows functors to be represented on a separate level and facilitates the discovery of patterns between them.

Note, however, that functors are distinctive by being acoustically less prominent than the elements around them. As we have seen, data from L1 acquisition suggests that this also makes them harder to track in spoken input: infants prefer listening to lexical over functional items in their native language (Shi & Werker, 2001; 2003), and have difficulties with the phonological encoding of function words (Hallé, Durand & de Boysson-Bardies, 2008; Shi et al. 2006; Shi & Lepage, 2008; Shi, Werker & Cutler, 2006), especially the less acoustically salient ones (Strömqvist, Ragnarsdóttir & Richthoff, 2001). It is possible that this lack of acoustic salience will make learners (adults or infants) less likely to focus on the target elements, or to detect the dependencies between them. A counter to Hypothesis 1, therefore, is that NAD-learning (at least in the linguistic domain) is reliant on the (perceptual) prominence (rather than the distinctiveness) of elements entertaining dependencies: the more prominent (i.e. higher in pitch, intensity, longer in duration) the elements, the easier it is to keep track of them (Strömqvist, Ragnarsdóttir & Richthoff, 2001), and therefore the easier it is to detect patterns between them. This predicts that dependencies between functors in natural languages would be difficult to detect because functors themselves are not perceptually prominent – implying that NAD-learning may not be a crucial mechanism supporting the acquisition of morpho-syntactic dependencies.

If Gestalt principles of perception are used to group functor-like units together and compute dependencies between them over lexical-like units, the next question is: what is the nature of the cues that are used to distinguish between different types of units. Shi, Morgan & Allopena (1998) as well as Monaghan, Christiansen & Chater (2007) combine a variety of cues from the linguistic (segmental phonology, syllable structure, etc.) as well as non-linguistic (acoustic) domain to mark the lexical/functional distinction. Functors are distinct from lexical words at a purely acoustic level (e.g. pitch, amplitude, etc.); it may be that these acoustic differences can also be exploited at a higher, linguistic level of analysis: prosody. Depending on the prosodic system of the language under study, functional words can be prosodic clitics (Selkirk, 1996), and therefore prosodically unmarked compared to their lexical counterparts, which can receive lexical stress or tonal accent.

Do Gestalt principles of perception operate at the primary level of acoustic perception, grouping together or dissociating elements based on their acoustic properties? Previous studies of NAD-learning with non-linguistic input have shown that NAD-learning is a domain-general mechanism, which can exploit a variety of acoustic cues: learners can detect dependencies between non-tonal over tonal noises (Gebhart, Newport & Aslin, 2009) or between high-pitched over low-pitched tones (Creel, Newport & Aslin, 2004). Thus, NAD-learning exploits a variety of acoustic cues in different types of input. One possibility, therefore, is that the domain-general NAD-learning mechanism exploits acoustic cues to group together non-adjacent elements in any type of auditory input (linguistic, musical, noise, etc.).

An alternative is that, with linguistic input, NAD-learning exploits more abstract levels of linguistic representation, such as segmental phonology (Newport & Aslin, 2004; Onnis et al., 2005) or prosody? The findings of Newport & Aslin (2004), Onnis et al. (2005) and Van den Bos et al. (2012) suggest that NAD-learning can, in principle, be guided by Gestalt principles at the level of linguistic/phonological analysis. If NAD-learning is to be a powerful tool for language acquisition, it should rely on linguistic representations available to the learner, which are more abstract and robust in the face of variation or noise conditions. Purely acoustic cues may guide NAD-learning of non-linguistic stimuli (Creel, Newport & Aslin, 2004; Gebhart, Newport & Aslin, 2009) but the acoustic factors identified by Shi, Morgan & Allopenna (1998) and Monaghan, Christiansen & Chater (2007) to distinguish between functional and lexical elements could at least in part correlate with the different prosodic properties that these two categories have. Therefore, **Hypothesis 2** states that if Gestalt principles of perception facilitate the detection of dependencies between functor-like over lexical-like elements, then these principles operate at the level of prosodic (as opposed to purely acoustic) representation of the input: prosodically reduced elements are grouped together and dissociated from prosodically unreduced elements.

Previous studies have looked into the role of prosody in NAD-learning. Peña et al. (2002) and Endress & Bonatti (2006) (as well as Marchetto & Bonatti, 2013, for 18 month-olds) showed that inserting subtle segmentation cues (25ms pauses) at the boundaries of *aXb* strings (as opposed to presenting them in a continuous stream) facilitated preference for 'rule-words' (novel *aX'b* strings with correct dependencies but novel intervening *X*s). Thus, prosodic cues to segmentation facilitated the generalization of dependencies. Langus et al. (2012) showed that phrase-final lengthening and intonational cues also facilitated a preference for rule-words, while Mueller, Bahlmann & Friederici (2010) showed that segmentation pauses as well as a rising-falling pitch contour over the relevant chunks facilitated the detection of center-embedded dependencies ($a_i a_j b_j b_i$). Prosody as a cue to relevant segmentation of the input therefore facilitates the detection of dependency relations. I propose that prosodic cues may support NAD-learning beyond their role in segmenting and organizing the input. The prosodic status of individual words can also facilitate the detection of dependencies between them.

2.5 Research Questions and Implementation

In this study I test two Hypotheses:

Hypothesis 1: The detection of dependencies between functor-like words over lexically-sounding words is facilitated by Gestalt principles of perception which group the perceptually similar functor-like units together.

Hypothesis 2: If Hypothesis 1 is correct, then Gestalt principles of perception operate at the level of prosodic representation of the input, grouping together prosodic 'clitics'.

To test Hypothesis 1, I examine the role of perceptual prominence vs. perceptual distinctiveness in a controlled AGL paradigm, by employing a simple artificial grammar (*aXb*, cf. Gómez 2002) where I vary the acoustic cues to the *a_b* dependencies across three different conditions (while maintaining the *X*s constant) and observe the effect on learning. Specifically, I test three Acoustic Conditions: an *Emphasized Condition*, where the a_i and b_i elements in the $a_i\_b_i$ dependencies are more acoustically prominent than the intervening Xs; a *Lexical Condition*, where the dependent elements, like the *X*s, have the perceptual properties of lexical items in natural languages (e.g. lexical stress, full vowels, etc.); and a *Functional Condition*, where the dependents have perceptual properties of functors (reduced vowel, lower pitch, intensity and shorter duration than lexical words), and are therefore less prominent than the *X*s. If Hypothesis 1 is correct, participants should learn the dependencies well in the Emphasized and Functional Conditions, but perform poorest in the Lexical Condition where the dependent elements are not perceptually distinctive. Conversely, if the dependencies' salience is a function of their acoustic prominence, performance will decline linearly over the three Conditions, and performance in the Functional Condition (where dependencies are least prominent) should be poorest. Under both hypotheses, performance in the Emphasized Condition should be superior to performance in the Lexical Condition. If perceptual cues do not affect NAD-learning, performance should not differ significantly across Conditions.

To disentangle whether learners exploit purely acoustic or prosodic cues, I construct stimuli in such a way that dependent elements (*a, b*) in the Emphasized and Functional Conditions are always marked by clear acoustic cues (pitch, duration and amplitude) that differentiate them from the intervening elements (*X*). The prosodic status of the words is determined by a combination of the acoustic cues (e.g. higher pitch indicative of prosodic markedness), and the inter-word pause cues. Gómez (2002) employed 250 ms pauses between words in an a_iXb_i string, with strings being read out in a lively, child-friendly voice emphasizing the *a/b* tokens. These are audible pauses that normally mark boundaries between prosodic units in natural speech and would therefore be appropriate for delimiting (acoustically/prosodically) marked items, such as the a/b tokens in the Emphasized Condition. However, functional-sounding elements often have the prosodic status of clitics (Selkirk, 1996), which means that they cannot be prosodic units of themselves but need to attach to a stem. I posit that shorter pauses, of 100 ms (which are near the auditory threshold for pause perception, cf. Zellner, 1994, but are long enough to eliminate the need for co-articulation at the word boundaries), would render the Functional Condition more natural.

I tested subjects in a 3x2 between-subjects design, where each of the three Acoustic Conditions was tested with 2 Pause Versions: one with 100 ms pauses between the words in an *aXb* string (as if the string was a single prosodic unit), and one with 250 ms pauses between words (as if the *aXb* string was composed of 3 self-standing prosodic units). Thus, while in the Emphasized Condition with 250 ms or the Functional Condition with 100 ms pauses, pause cues were consistent with acoustic cues in marking the prosodic status of the dependent *a/b* elements (as either accented or, respectively, reduced), the Emphasized Condition with the 100 ms pauses, as well as the Functional Condition with 250 ms pauses contained conflicting cues as to the prosodic status of *a/b*. For instance, in the Functional Condition with 250 ms pauses, *a/b* elements were recorded as prosodic clitics, but they were separated by the stem they should attach to by audible pauses.

If learners only relied on acoustic cues to mark the dependent elements *a/b* as similar to each other and distinct from the context, pause cues should be irrelevant to their learning performance and only an effect of Acoustic Condition should obtain. If, on the other hand, learners employ Gestalt principles of perception on the prosodic level, then learning performance should decline in the conditions where acoustic vs. pause cues are inconsistent in marking the prosodic status of the *a/b* elements. Namely, with functional-sounding *a/b* tokens, a shorter intra-stimulus pause should promote learning, since functors are often prosodic clitics (Selkirk, 1996); with highly emphasized *a/b* tokens, a longer pause should serve to mark their salience. In this case, I expect a significant Acoustic Condition by Pause Version interaction, and better performance with 250 ms rather than 100 ms pauses for the Emphasized condition, and the reverse pattern for the Functional Condition.

2.6 Experiment 1

I adopted the design of Gómez (2002) with stimuli (adapted to Dutch participants) from Grama et al. (2013). I tested NAD-learning in the three different Conditions described above: Emphasized, Lexical and Functional, each with two Pause versions, 100 ms or 250 ms. Therefore, the design was a 3 (Acoustic Conditions) x 2 (Pause versions) full-factorial, between-subjects design. The methodology was the same across conditions: as in Gómez (2002), participants were exposed to a language consisting of three *a_b* dependencies combined exhaustively with a set of *X* elements. Following this familiarization, they were tested on their knowledge of the dependencies by receiving *aXb* strings either with correct, or incorrect dependencies and having to indicate, for each, whether they thought it was consistent with the language they had heard.

I introduced two important modifications to the methodology in Gómez (2002). Firstly, Gómez (2002) employed *aXb* test strings where the intervening *X* element was taken from the familiarization phase: thus, the correct test items, where the dependency was consistent with familiarization, were in fact *aXb* strings which had been heard in the familiarization. Previous literature has shown a distinction between the ability to recall chunks previously heard, and the ability to learn a given pattern as a rule, that is, to be able to use it productively and generalize it to novel contexts (Peña et al., 2002; Endress & Bonatti, 2006; Endress & Mehler, 2009, among others). In this study I am concerned with learner's ability to obtain knowledge of non-adjacent dependencies as generalizable rules, and to be sensitive to these rules even when they are instantiated in unfamiliar contexts. This is the type of ability which will serve language acquisition in aiding learners to detect grammatical patterns as generalizable/productive rules of grammar, and not as patterns that occur in familiar contexts. In order to ensure that participants are not recalling chunks from familiarization but generalizing dependencies to novel contexts I employ *aX'b* test strings where the intervening *X'* element has not occurred during familiarization, ensuring that all test items were novel.

Secondly, in Gómez (2002) participants were familiarized with the artificial language with the explicit instruction that they should listen carefully because they would be subsequently tested on their knowledge of this language. This means participants were faced with a single task of listening intently, and could explicitly look for regularities that might improve their chances of success at test. However, in this study I aim to investigate a learning mechanism that may possibly aid (L1) language acquisition. It is a general consensus in the literature that infants acquire their native language under incidental learning conditions: they do not benefit from explicit instruction, presumably direct their attention to understanding the meaning rather than acquiring the structure of the input, and acquire language in an environment where various other stimuli or tasks may distract their attention (see Saffran et al., 1997, and references therein). Incidental learning in adults has been shown to closely resemble incidental learning in children, in a setup where participants were deterred from explicitly focusing on the spoken input by performing a simultaneous task of coloring (Saffran et al., 1997). Introducing a simultaneous task during an artificial grammar learning experiment has been shown to generally affect explicit, but not incidental learning (see Dienes & Berry, 1997 and references therein), especially where the secondary task is in a different modality, or engages different computational mechanisms than the first. In short, adult incidental learning, induced by introducing a secondary, unrelated task, is likely to be a good model for a learning mechanism which might subserve early language acquisition. In this study I employ a coloring task, similar to Saffran et al. (1997).

2.6.1 Participants

A group of 149 adult monolingual Dutch participants (17 male, age range 18-47, mean age 22) were recruited from the Utrecht Institute of Linguistics OTS database for adult participants and paid 5 euros for participation. A majority of them were students at the Utrecht University, in other areas than linguistics. Participants were required to have no hearing impairment and no diagnosis of dyslexia or attention deficits. One participant (female) was excluded (from the Functional Condition with 250 ms pauses) for having previously participated in a similar experiment.

2.6.2 Materials

Familiarization. In all three Conditions, two *aXb* (e.g. *tep naspu lut*) languages L1 and L2 were created (same as in Gómez, 2002), each with three $a_i\_b_i$ pairs (100% conditional probability); language L1 contained the pairs *tep_lut*, *sot_jik* and *rak_toef*, whereas language L2 contained the pairs *tep_jik*, *sot_toef* and *rak_lut*, such that every $a_i\_b_i$ pair in one language was ungrammatical ($*a_i\_b_j$) in the other. Subjects were randomly assigned to L1 or L2 to avoid stimulus effects (e.g. if for some reason *tep* is combined more easily with *lut* than *jik*).

A set of 18 different bisyllabic *X* words was used (see Appendix). The *a*, *X* and *b* tokens were recorded separately (see below) and combined into *aXb* strings with either 100 ms or 250 ms pauses between the words (*a* and *X*, and *X* and *b*), depending on the Pause condition. During the familiarization, participants heard 324 strings (3 *a_b* pairs x 18 *X*s x 6 repetitions, randomized per participant), with 750ms pauses between each two strings. I chose set size 18 for the intervening *X* elements because it has been shown to work just as well as set size 24 with infants (Gómez & Maye, 2005), and I wanted to test if the same could be said for adults. If this set size is large enough to yield learning in some conditions but not others, it will highlight the effect of the perceptual cue manipulations that I wish to investigate. I chose bisyllabic intervening elements in line with Gómez (2002) to further facilitate the distinction between *a/b* tokens and *X*s. As said before, NADs have been shown to only be learned in cued contexts (Peña et al., 2002; Onnis et al., 2005), and in this experiment I cued dependent *a/b* elements both in terms of their high frequency/low variability and their (syllabic) length. If there is no evidence of learning in some conditions, this must be due to the specific perceptual manipulations introduced.

Test. Two *X*s (novel, ensuring that participants could generalize the dependency rules), were combined with the three $a_i\_b_i$ pairs of L1 (ungrammatical for learners of L2), and the three $a_i\_b_i$ pairs of L2 (ungrammatical for L1), for a total of 12 test strings (6 of which were consistent with L1, and 6 with L2).

Stimuli for the Lexical and Functional Conditions were recorded in a sound-attenuated booth, at a sample frequency of 48 kHz, using a TASCAM DA-40 DAT-recorder. A female native Dutch speaker read out sentences in Dutch, each containing a nonce word, as naturally as possible. All *X*s, as well as Lexical *a/b*s, were recorded in the syntactic slot where a direct object noun would normally be found, in the template sentence:

(4) *Ik zie de ____ in de tuin.*

I see the __ in the garden.

e.g. *Ik zie de fapoeg in de tuin.*

Ik zie de tep in de tuin

The *a* and *b* elements in the Functional Condition were recorded in the same carrier sentences, except they now filled the position of the determiner preceding the direct object (5). In all instances the speaker was instructed to realize the nonce words in accordance with their syntactic position:

(5) *Ik zie __ aapje in de tuin.*
I see __ monkey in the garden
e.g. *Ik zie tep aapje in de tuin.*

For the Emphasized Condition, *a* and *b* elements were taken from Grama et al. (2013). The same female reader read out strings of four nonsense words (two intonational phrases, e.g. [*lotup tep*] [*poemer lut*]), in a lively manner, with emphasis on the monosyllabic words. Note that the methodology for recording stimuli for the Emphasized Condition is highly similar to that employed by Gómez (2002) for obtaining her stimuli (reading out the nonsense strings with lively, child-friendly intonation rendered the *a/b* elements in the string highly acoustically salient): the Emphasized Condition with 250 ms within-string pauses was therefore designed to approach Gómez (2002) as closely as possible (with the exception of the methodological changes introduced, namely the secondary task designed to elicit implicit learning and the use of novel test items to test rule-learning).

Acoustic measures, performed in Praat 5.3.03 (32-bit Edition for Windows, Boersma & Weenink, 2011), of the *a/b* tokens for the three Acoustic Conditions are presented in Table A2.1 of the Appendix, and show clear differences in acoustic properties between the three Conditions. Acoustic measures for the X elements employed for familiarization and test are presented in Table A2.2 of the Appendix.

With these stimuli I ran a validation experiment in which I asked 9 naive listeners to match *aX/bX* pairs, containing either lexical or functional *a* and *b* tokens, with real Dutch noun phrases, composed of either an adjective (lexical) and a noun or a determiner (functional) and a noun (on the basis of acoustic/prosodic similarity). For instance, in a two-alternative forced-choice task the participants would be given an *aX* pair in an 'alien' language (e.g. *tep poemer*) where the *a* sounded either functional or lexical, and would be asked to 'translate' this alien phrase into either the phrase *een tijdschrift* (a magazine, determiner+noun) or the phrase *oud tijdschrift* (old magazine, adjective+noun), respectively. Participants correctly assigned experimental *aX/bX* pairs containing 'functional' or 'lexical' *a/b* tokens to the targeted categories (determiner+noun or adjective+ noun), in 79.6% of the cases (SD=22.49). A one-sample *t*-test confirmed that the accuracy score of 79.6% was significantly above chance, $p = .004$. Thus, the artificial stimuli resembled to a fair degree the perceptual properties of Dutch functional and lexical elements.

2.6.3 Procedure

Familiarization. Participants were seated in a sound-attenuated booth, coloring a mandala while listening to an 'alien language'. They were instructed to 'listen passively' and attend primarily to the coloring. To avoid any motivation to explicitly look for patterns in the stimuli, participants were not informed of the subsequent test phase. The familiarization phase lasted between 10 and 15 minutes, depending on the Acoustic Condition and Pause version, and consisted of 324 aXb strings played out in a randomized order with 750ms silences in between.

Test phase. After the familiarization, participants were told that the language they had heard had certain regularities related to word order and that they would hear 12 new sentences in this language, only six of which conformed to its rules. They would have to give grammaticality judgments for each of the strings based on their intuition. The test strings were presented in random order one after the other, and while each string was played, a question appeared on the screen asking participants: Does this sentence belong to the language you have just listened to? Note that the test strings had the same perceptual properties as the familiarization strings in each experiment, as the a/b tokens were identical to familiarization, and (novel) X' tokens were recorded in the same way/session as the familiarization Xs. After hearing the test string participants responded 'yes' or 'no' by pressing one of two buttons on a button-box.

After the test, participants were debriefed on what they had noticed about the language they had heard, and what strategies they had used in answering the questions, if any. They were also asked to rate, on a scale from 1 to 7, their confidence in the responses they had given at test. According to the zero-correlation criterion (see Dienes, 2007, and references cited therein) participants are implicit learners if their assessment of their own performance does not correlate with their actual performance. I wanted to see how implicit or explicit participants' knowledge of the structure of the strings was: if participants who performed better on the test also expressed higher confidence in their answers, then there was some explicit awareness of the existence of structure in the input.

2.6.4 Results

To assess learning performance in each of the 6 Conditions (3 Acoustic x 2 Pause), I ran one-sample *t*-tests on the mean Accuracy scores (percentage correct responses per participant) for each of the 6 cells, comparing each to chance (see Table C2.1 and Figure C2.1). Participants in the Emphasized Condition with 250 ms pauses performed significantly above chance (t (24) = 3.674, p = .001), whereas learning did not reliably differ from chance expectation in any of the other 5 Conditions.

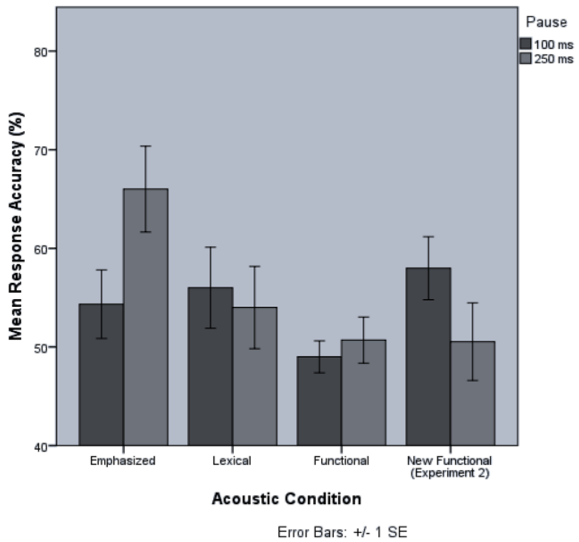

Figure C2.1 Mean Accuracy scores per Acoustic/Pause Condition for Experiments 1-2

To compare performance across the 6 Conditions, I ran a Generalized Linear Mixed Model analysis (using IBM SPSS version 20.0.0), with Accuracy (correct responses, meaning correct rejections of ungrammatical, and correct acceptance of grammatical test strings) as a (binomial) dependent variable. I introduced Subjects as a random factor, and Acoustic Condition (Emphasized, Lexical, Functional), Pause version (100 ms, 250 ms), and the interaction Acoustic Condition x Pause version as fixed factors. I also introduced Language (L1, L2) as a fixed factor, to control for stimulus-specific biases (the possibility that certain a_b combinations were inherently easier to learn than others), and the interaction Acoustic Condition x Language (because I used the same *a/b* words but with different perceptual properties in each Condition, I wanted to control for the possibility that the manner of recording of these different stimuli had introduced a bias for certain *a_b* combinations in some but not all of the Acoustic Conditions).

		EMPHASIZED	LEXICAL	FUNCTIONAL	New FUNCTIONAL (Experiment 2)
250 ms	Sample size	N = 25	N = 25	N=24	N=24
	Mean accuracy	66% (SD=21.8)	54% (SD=20.9)	51% (SD=11.5)	49% (SD=16.6)
		***p* = .001** 95% CI [7, 25]	***p*** = .347 95% CI [-4.6, 12.6]	***p*** = .770 95% CI [-4.2, 5.6]	***p*** = .680 95% CI [-8.6, 5.4]
	Effect size	***d* = .750**	***d* = .194**	***d* < .001**	***d* = -.087**
	Explicit learners	***4***	***3***	***0***	*7*
100 ms	Sample size	N = 24	N = 25	N=25	N=23
	Mean accuracy	54% (SD=17.54)	56% (SD=20.5)	49% (SD=8.09)	58% (SD=15.63)
		p = .297 95% CI [-3.6, 11.2]	***p*** = .156 95% CI [-2.5, 14.5]	***p*** = .543 95% CI [-4.3, 2.3]	***p* = .017** 95% CI [4.4, 4.5]
	Effect size	***d* = .218**	***d* = .371**	***d* = .060**	***d* = .516**
	Explicit learners	***9***	***4***	***0***	***3***
	Total mean accuracy	**60% (SD=20.55)**	**55% (SD=20.5)**	**50% (SD=9.84)**	**53 % (SD=16.63)**

Table C2.1. Results for Experiments 1-2 per Acoustic/Pause Condition, with mean accuracy rates, *p*-values for one-sample *t*- tests comparing mean accuracy rates to chance (and nonparametric, one-sample Wilcoxon Signed-Rank test for the New Functional Condition with 100 ms pauses), and effect size as Cohen's *d*.

There was a significant effect of Acoustic Condition (F (2, 1.767) = 4.161, p = .016), with Bonferroni planned comparisons yielding a near-significant difference between the Emphasized and Functional Conditions (t (1) = 1.884, p = .06, 95% CI [-2, 9.9]) but no other main effects or interactions (no effect of Language, p = .914, or Pause, p = .185, and no Acoustic x Pause, p = .115 or Acoustic x Language, p = .213 interaction). The difference in performance between the Emphasized and Functional Condition, therefore, approached significance. The Lexical Condition showed accuracy rates in between those of the Emphasized and Functional Conditions, respectively, and not significantly different from either. This pattern of results suggests that NAD-learning decreased with the decrease in acoustic prominence of the dependent elements.

2.6.5 Discussion

Experiment 1 tested learning of NADs in three different Acoustic Conditions and two different Pause versions; intervening *X* elements were kept the same throughout Conditions, while the perceptual properties of dependent elements *a* and *b* in *aXb* strings were varied (ranging from Emphasized, Lexical-sounding and Functional-sounding). There was a significant effect of Acoustic Condition, reflecting a marginally significant improvement in performance in the Emphasized Condition with respect to the Functional Condition. This suggests that adult participants were influenced by the acoustic prominence of the dependent tokens, such that the more perceptually prominent the elements, the easier the dependencies were to learn. Only the Emphasized Condition with 250 ms pauses, which emulated the design of Gómez (2002), yielded learning that was significantly above chance. Note that the learning effect in this condition was lower than the general learning performance in Gómez (2002). As mentioned above, in the current study I changed important aspects of the methodology, by promoting incidental over explicit learning, and testing participants' ability to generalize these dependencies to novel strings (with unfamiliar intervening Xs). These changes were implemented to more realistically reflect the natural language learning situations I am trying to model, but they may also have rendered the task more difficult.

The fact that there was significant learning in the Emphasized Condition only in the 250 ms Pause Version, and not in the 100 ms Version may suggest that participants found it easier to exploit the perceptual cues in the condition with the more naturalistic prosody. However, the analysis of the data did not yield a significant Acoustic Condition * Pause Version interaction, therefore there is no basis to draw a conclusion about the role of prosody in NAD-learning.

Verbal reports completed after the experiment revealed that some participants in the Functional Conditions did not segment the aXb strings as intended. Nine participants in the 250 ms version and four in the 100 ms one reported that they had perceived familiarization strings as having the syllable structure 1 – 1 – 2 (as opposed to the correct 1 – 2 – 1), suggesting the possibility that they had segmented the string-final element b as the initial element of the subsequent string (*baX*, with a 1 – 1 – 2 syllable structure, as opposed to *aXb*). Participants in the Lexical and Emphasized Conditions never reported this segmentation percept: the participants in those conditions who recalled the structure of the strings unanimously reported the correct 1 – 2 – 1 structure. While overall endorsement scores for participants in the Emphasized and Lexical Conditions showed a significant bias towards accepting strings as correct (mean endorsement rate 65.5%, SD=16.96, significantly above 50%, t (98)=3.661, p <.001), participants in the Functional Conditions exhibited no such bias (mean endorsement rate 46.09%, SD=26.9, t (48)=-1.018, p =.314); seven participants in the Functional Conditions rejected all strings in the test phase, while none did so in any of the other Conditions. Thus, while participants in the Emphasized and Lexical Conditions were biased to accept more test strings (partly) because of their structural similarity with the familiarization strings, participants in the Functional Conditions showed no such bias, suggesting that at least some of them were mis-segmenting familiarization strings and finding it difficult to endorse the differently structured test strings.

This mis-segmentation introduced a confound in the results of the two Functional Conditions that renders the results unreliable. I attributed mis-segmentation to the prosodic properties of the *b* elements: because these elements were recorded as phrase-initial nonce determiners, their prosodic contour was not appropriate for a string-final position. Thus, the prosodic contour of the *aXb* strings was unnatural, or rather the pause segmentation cues conflicted with the prosodic segmentation cues. As a consequence, some participants ignored the long, 750ms pauses between strings (that separated a *b* token from a subsequent *a* token), and combined the last word from a string to the first two words from the next string in a single prosodic phrase (*baX*). I wanted to eliminate the possibility that participants in the Functional Conditions performed poorly solely because of the unnatural prosodic contour of the familiarization stimuli, and its interference with the correct segmentation of the strings. For this purpose, I ran a new version of the Functional Conditions, in which the prosodic contour of the *aXb* strings facilitated their segmentation and did not conflict with the pauses that delimited them.

2.7 Experiment 2

2.7.1 Participants

I recruited 51 participants (5 male, age range 18-42, mean age 22) in the same way as before; 4 participants were excluded, 3 due to technical problems and one for familiarity with research on NAD-learning. Of the remaining 47, 24 were assigned to the New Functional Condition with 100 ms pauses, and 23 to the New Functional Condition with 250 ms pauses. Subjects were again randomly assigned to one of the languages L1 or L2.

2.7.2 Materials

Tokens of the artificial grammar were re-recorded in similar fashion as before: the *X* items were recorded in the same carrier-sentence (4). For the *a* and *b* tokens, I chose the morpho-syntactic dependency between the neuter determiner *het* and the diminutive suffix *–(t)je* in Dutch as a model, and recorded the *a* and *b* tokens as the determiner and suffix respectively:

(6) *Ik zie het zebra'tje.*

I see the zebra.DIM

e.g. *Ik zie tep zebra'tje*

Ik zie het zebralut.

I analyzed the *a/b* tokens acoustically, the same way as before (see Table A2.3 in the Appendix). The testing procedure was identical to Experiment 1.

2.7.3 Results

I compared the results in each Pause version of the New Functional Condition with chance (50%) performance: a one-sample *t*-test on the mean Accuracy scores (percentage correct responses per participant) for the New Functional Condition with 250 ms pauses revealed no significant learning effect (p = .680, 95% CI [-8. 6, 5.4], Cohen's d = -.087), with participants clearly scoring at chance (M=48.55%, SD=16.6). The mean Accuracy score in the New Functional Condition with 100 ms pauses was M=57.99% (SD=15.63). Because a one-sample Kolmogorov-Smirnov test showed that the Accuracy scores were not normally distributed (p = .043), I ran a non-parametric test (one-sample Wilcoxon Signed-Rank test) on the mean Accuracy scores, which showed that performance was significantly better than chance (Median = 0.58, SE = 18.78, Z = 2.396, p = .017) with a Cohen's d effect size of .516 (see Table C2.1).

I compared the new Functional Conditions (100 ms and 250 ms) with the old Emphasized and Lexical Conditions in a Generalized Linear Mixed Model as before (Acoustic Condition, Pause, Language, Acoustic x Pause and Acoustic x Language as fixed factors, and Subject as random factor). There was no significant effect of Pause (p = .921), Language (p = .487), Acoustic Condition (p = .175), or Acoustic Condition x Language interaction (p = .236), but there was a significant Pause x Acoustic Condition interaction (F (2, 1.743) = 3.819, p = .022). Whereas participants in the Emphasized Condition showed significant learning in the 250 ms Pause Version but not the 100 ms Pause Version, the Functional Condition showed above-chance performance in the 100 ms Pause Version but not the 250 ms one, suggesting that learning was promoted when the pause cues *together* with the acoustic cues (pitch, duration, etc.) marked the dependent elements as prosodically distinct from the intervening material.

Contrary to the Functional Conditions of Experiment 1 (and similar to the Emphasized and Lexical Conditions), participants in Experiment 2 showed a significant bias towards accepting test items as correct (overall endorsement rate M=62.94%, SD=19.26, t (46) = 4.608, p <.001), suggesting that they were sensitive to the structural similarities between familiarization and test strings. Only one participant in 47 rejected all test strings, and none reported an incorrect segmentation strategy. Together with the improvement in learning in the New Functional Condition with 100 ms pauses this suggests that mis-segmentation was no longer an impediment to learning in Experiment 2.

I also pooled together the data from Experiments 1 and 2 and ran a two-tailed Pearson correlation test to verify whether participants' confidence in their responses correlated with their actual performance. I obtained a significant correlation between the confidence ratings on a 1-7 scale and the accuracy rates per participant (.198, p = .006). Some participants reported being aware of the presence of a dependency between the first and last word of the strings: Table C2.1 shows, for each condition, how many participants reported awareness of a dependency ('Explicit learners'). However, when I excluded these participants who were explicitly aware of a pattern, the correlation between performance and confidence ratings was non-significant, both in the overall dataset (-.068, p=.385) and in the individual conditions, suggesting that the initial effect was carried exclusively by the high confidence ratings of the explicit learners. None of the participants that became explicitly aware of the pattern reported intentionally looking for patterns in the input. Instead, all of them reported that at some point in the familiarization phase they suddenly became aware of dependencies, due to the frequent occurrence of the short (dependent) nonce-words. In general, awareness of the properties of the familiarization language ranged from not being able to indicate the number of words in a strings, to being able to reproduce all three dependencies, or being aware of the limited set of first and last words (*a/b*) that combined productively with the larger set of intervening *X*s, without reporting the existence of dependencies between the first and the last words. All participants colored a substantial part of the mandala, indicating that they were committed to the coloring task throughout (at least a large part of) the familiarization phase.

2.7.4 Discussion

I retested participants' learning of NADs in the Functional Condition with familiarization strings exhibiting a rising-falling pitch contour: I re-recorded the stimuli in the same way as before, with the minor difference that the b elements were recorded as phrase-final functors in a natural morpho-syntactic template. This allowed the *aXb* strings in the familiarization to have a rising-falling pitch contour similar to prosodic phrases in natural languages, and eliminated the risk of mis-segmentation (see Figure A2.1 in the Appendix).

Performance in the Functional Conditions improved selectively: participants' accuracy in judging the test items was significantly above chance when intra-stimulus pauses were 100 ms, but were at chance when the pauses were 250 ms. By contrast, participants in the Emphasized Conditions performed above chance with 250 ms pauses, but not with 100 ms pauses. This result is not unexpected. I assumed that if the within-string pauses played a role, the 100 ms pauses would facilitate learning in the Functional Condition, because the functional-sounding *a/b* tokens had the prosodic status of clitics (Selkirk, 1996), and would therefore sound more natural if isolated by shorter pauses ; I also assumed that 250 ms within-string pauses would be more appropriate for the stimuli in the Emphasized Condition, as in the latter the *a/b* tokens had the status of highly emphasized words, which in natural speech can often be separated by a perceivable pause from the rest of the sentence.

The above-chance performance in the Emphasized 250 ms and the New Functional 100 ms Conditions suggests that NAD-learning is optimal (consistent with Hypothesis 2) when the dependent elements are different in salience from the intervening material, but integrated into a prosodically natural contour. The significant Pause by Acoustic Condition interaction suggests that prosody is crucial to NAD-learning: learners employ the acoustic cues in the input to establish the prosodic status of each element in a string, and then apply Gestalt principles at the prosodic level of organization to aid the computation of co-occurrence statistics between prosodically (rather than acoustically) similar elements. If prosodic cues are conflicting (pauses separating words are not consistent with the prosodic status of the word), the computation is rendered more difficult. The results for the retested Functional Condition contradict the initial interpretation of Experiment 1: learners are able to detect dependencies even when the dependent elements are less salient, provided they are prosodically distinct from the intervening material. These results are in line with the hypothesis put forth by Newport & Aslin (2004) that NADs are detected through a mechanism of Gestalt perception, with the amendment that for linguistic input Gestalt principles are applied not (only) directly at the acoustic level, but at the higher, language-specific level of prosody.

It is important to note that in Experiments 1 and 2 participants reached varying levels of awareness: from failing to indicate the correct number of words in a string (3), to indicating the syllabic structure of an *aXb* string (1-2-1), to indicating some/all of the *a/b* words and their position in a string, and finally to identifying the presence of a dependency between the first and last word in an *aXb* string and being able to recall some/all of the dependencies. One suggested measure to assess whether there is implicit learning (Dienes, 2007) is to check for correlations between participants' assessment of their performance, and their actual performance. If this correlation exists, then participants have at least some explicit awareness that their answers are being guided by knowledge of the structure of the language. When I excluded participants aware of the rules of the language, I found no correlations between participants' accuracy rate and their overall confidence in their performance, and therefore no evidence of explicit awareness.

One pertinent question that arises is whether the participants who became aware of the dependencies represent a distinct group from those who didn't. In other words, does explicit knowledge represent a confound? I would argue that this is not the case. Firstly, all participants received the same instructions, and all explicit learners reported that they had obeyed the instructions, listening to the language passively and only noticing the dependencies (after hearing a substantial amount of input) due to their recurrence in different strings. All participants were exposed to the language in incidental learning conditions, unaware that they had to acquire the structure of the language, or what type of structure it was. All were given a simultaneous task and none were warned about the existence of a subsequent test phase, so as to discourage memorizing the input or looking for patterns. Therefore, there is nothing to indicate that the procedure differed across participants. Why then would the outcome, namely the amount of implicit/explicit knowledge derived from the same incidental learning process differ per participant? The literature (see Reber, 1989, and references cited therein) emphasizes that implicit learning is more likely to obtain with complex grammar, made up of multiple rules. In this study I tested the acquisition of a simple rule: the dependency between the first and last words in a string, and thus allowed for the possibility that participants would derive explicit knowledge of that rule if their sensitivity to it exceeded a certain threshold. Note that the distribution of explicit learners is not even across conditions, but there are more explicit learners in the conditions predicted to induce above-chance performance (9 in the Emphasized Condition with 250 ms pauses and 7 in the New Functional Condition with 100 ms pauses) than in the rest of the conditions (3-4). Instead of assuming that participants in these specific conditions happened to be less likely to follow the instructions, I propose that it is more likely that these condition promoted explicit awareness of the rules by facilitating the detection of those rules.

The results of Experiment 2 are consistent with the hypothesis put forth by Newport & Aslin (2004) and subsequent studies, that NADs are detected through a mechanism of Gestalt perception, whereby elements that are prosodically distinctive are grouped together, and patterns between them are thus more easily learned. However, an important aspect of Newport & Aslin's account is the fact that the prosodic similarity between the dependent elements is crucial for the detection of dependencies. In the experiments above I showed that the prosodic distinctiveness of these elements plays a definite role, however, the role of similarity between the dependent elements *a/b* was not directly assessed.

One possibility is that dependent elements do not have to be similar to each other , but only have to be different from the intervening material. Thus, prosodic cues may simply need to mark a distinction between *a*/*b* elements and the intervening *X* (as opposed to the Lexical Condition in Experiment 1 where all three elements had the prosodic status of lexical words), but *a* and *b* may not share prosodic properties. In order to eliminate this possibility, Experiment 3 was set up to investigate learning between perceptually distinctive, but dissimilar elements. I employed exactly the same design and stimuli as above, but tested participant's NAD-learning in *aXb* strings where the *a* element was Emphasized, the *X* elements was Lexical-sounding and the *b* element was Functional-sounding.

2.8 Experiment 3

I took stimuli from the Emphasized and New Functional Conditions, and combined them into new *aXb* strings where the *a* was Emphasized (with 250 ms pauses between *a* and *X*) and the *b* was Functional (with 100 ms pauses between *X* and *b*, and a phrase-final falling pitch contour on *b*). Thus, the structure of the strings was Emphasized – Lexical – Functional, with the dependent elements being different in perceptual properties both from the intervening *X* and from each other. If participants could learn such dependencies as well as they learned them in a Functional – Lexical – Functional configuration, or in an Emphasized – Lexical – Emphasized one, then what promotes NAD-learning is not necessarily a principle of similarity between dependent elements, but simply their ability to 'stand out' in the speech stream.

2.8.1 Stimuli

I used the *aXb* strings from the New Functional Condition (with 100 ms) in Experiment 2; I spliced the Functional *a* elements, and replaced them with Emphasized *a* elements from Experiment 1, and made the pauses between *a* and *X* 250 ms, so as to be faithful to the prosodic naturalness that proved so important to learning in Experiments 1 and 2. The resulting strings therefore had the structure Emphasized *a* – 250 ms pause – Lexical *X* – 100 ms pause – Functional *b*, with falling pitch contour on the final *b*. Participants were again each randomly assigned to one of the languages L1 or L2.

2.8.2 Participants

I recruited 25 participants (4 Male, mean age =21.04 years). One participant (female) was excluded for being familiar with research on NAD-learning and detecting the rules of the language based on this knowledge. The testing procedure was identical to the previous two experiments.

2.8.3 Results

The mean accuracy score reflecting the percentage of correct responses was M=55.56% (SD=20.66). A one-sample t-test on these scores showed that participants did not perform significantly above chance (*t* (23)=1.318, *p* =.201, 95% CI [-3.2, 14.3]).

2.8.4 Discussion

In this last experiment I asked whether the perceptual distinctiveness (from *X*) of dependent elements (*a/b*) was sufficient to highlight NADs for learning, without the necessity of there being perceptual similarity between the two. I found no evidence of above-chance performance in detecting *a_b* dependencies when the *a* and *b* elements were perceptually distinctive (with respect to the intervening *X*), but differed in perceptual properties from each other (one was Emphasized and one was Functional, while the intervening *X* was Lexical). One possible confound, however, was the use of different lengths of pauses between different words in an *aXb* string: 250 ms pauses between *a* and *X*, and 100 ms pauses between *X* and *b*. Could this difference have affected the segmentation, or grouping of elements, in the strings? There were no reports of erroneous segmentation: participants who reported on the familiarization strings reported the correct 1-2-1 structure. Subjects also showed a significant bias towards endorsing test strings (overall endorsement rate M=63.89%, SD=17.14, *t* (23) = 4.969, *p* =.001), suggesting their sensitivity to the structural similarities between familiarization and test strings. Thus, there is no reason to believe the non-uniform nature of the strings in this third Experiment interfered with participants' perception of the strings – rather that the perceptual difference between the dependent elements did not support the detection of dependencies between them.

This result is in line with Hypotheses 1 and 2 that NAD-learning is guided by Gestalt principles of perception which require dependent elements to be both prosodically similar to each other *and* distinct from the surrounding material. However, it is important to note that a null result in itself is not evidence for or against a hypothesis. With slightly different stimuli (e.g. a Functional *a* – Lexical *X* – Emphasized *b*), or a larger sample of participants, a significant learning effect might have emerged. Further investigation is needed to establish whether perceptual similarity between dependent elements is crucial to the detection of dependencies, bearing in mind that perceptual/prosodic difference between dependent elements may also arise in natural languages. For instance, in a language like Spanish most functional words do not possess lexical stress, with the exception of pronouns and auxiliary verbs. In a sentence like *¿Cuanto tiempo <u>has</u> trabaj<u>ado</u>?* ('How much time <u>have.2PL</u> work<u>ed</u>'), therefore, the auxiliary is metrically prominent, while the participle suffix is not, since lexical stress is assigned to the penultimate syllable in Spanish words ending in a vowel. Thus, while the free functor is prosodically strong, the bound morpheme is prosodically weak. Similarly, while a language like French does not have lexical stress, it does assign prosodic prominence to a phrase-final, full-vowel syllable (Jun & Fougeron, 2000). Therefore, a dependency like the one in *Vous chantez* (You.pl sing.2ndpl) is instantiated between a non-prominent pronoun and a prosodically prominent agreement suffix. The existence of such examples may raise the question of how productive a learning mechanism based on Gestalt principles of perception would be cross-linguistically, and how dependencies such as the one exemplified here could be learned. Would speakers of Spanish or French be better able to detect dependencies between a prosodically reduced and a prosodically prominent element than Dutch speakers? The findings of Experiment 3 need to be extended by employing different types of stimuli, as well as participants with different L1 backgrounds, in order to identify the role of Gestalt principles of perception above and beyond the processing and learning biases that adult participants might derive from their native language.

It is important to note, however, that Gestalt perception operating at the level of prosodic representation is just one of the different types of cues that may support NAD-learning. In these experiments I investigated the possibility that Gestalt principles of perception may aid the detection of dependencies that are instantiated between perceptually non-salient elements, since lack of salience has been hypothesized to inhibit tracking of these elements in the input (Shi & Lepage, 2008; Stromqvist, Ragnarsdóttir & Richthoff, 2001). Therefore, I set out to see whether Gestalt perception may compensate for the lack of salience of these elements; in languages like Spanish or French, however, functors may be more perceptually salient than in a language like English (Shi, Cutler, Werker & Cruickshank, 2006), and therefore dependencies between them may be more easily tracked simply because the dependent elements themselves are more easily tracked. Briefly put, languages like Spanish or French may not need to rely on Gestalt principles of perception, as other cues would be more readily available to aid the NAD-learning process. While it is beyond the scope of this dissertation to exhaustively explore how Gestalt principles of perception would aid or inhibit the detection of morpho-syntactic dependencies across languages with different prosodic properties, a fruitful avenue for future research would be the identification of specific cues (whether phonological, prosodic or distributional) across languages with different prosodic properties, that might support detection of morpho-syntactic dependencies.

2.9 General Discussion

In this study I investigated the role of perceptual factors in the process of learning dependencies between non-adjacent elements in spoken input. I asked whether perceptual factors can influence a distributional learning mechanism like NAD-learning at all, and if so, how perceptual cues might affect NAD-learning. Participants' ability to learn dependencies between non-adjacent elements in a simplified artificial language (*aXb*) was tested by measuring their endorsements of novel strings (with a novel intervening *X'*) with either grammatical ($a_iX'b_i$) or ungrammatical ($a_iX'b_j$) dependencies. Participants' learning performance, quantified as their endorsement of correct dependencies and rejection of incorrect ones, was modulated by the differences in acoustic/prosodic properties of the dependent elements. The results of Experiment 1 suggested that NAD-learning performance declined when perceptual cues rendered dependent elements less prominent: participants acquired dependencies between highly acoustically salient words (Emphasized Condition), but were not sensitive to these dependencies when the target words were not particularly salient (Lexical Condition), or when the target words were phonetically 'reduced' (Functional Condition). However, performance in the Functional Condition may have been affected by a confounding factor, namely the unnatural prosodic contour of the stimuli leading to an erroneous segmentation strategy. When I eliminated this confound, the picture I obtained was quite different. Participants were successful in detecting dependencies both between highly prominent (Emphasized), and between perceptually 'reduced' elements (Functional), but only when the strings had specific prosodic properties: there was reliable discrimination of grammatical and ungrammatical strings in the Emphasized Condition only in the 250 ms Pause version, and in the Functional Condition only in the 100 ms version.

It is important to note that these conclusions rely on a comparison between the findings in Experiments 1 and 2. Stimuli for Experiment 2 were recorded afresh, under the same conditions and with the same speaker instructions so that they matched the stimuli in Experiment 1 as closely as possible apart from the necessary manipulations. Furthermore, the nature of the task in these experiments demands a between-subjects design. Thus, Experiments 1 and 2 differed in stimuli, participants, and time of data collection, despite my best efforts to minimize those differences (by ensuring the stimuli were comparable in acoustic properties, by recruiting and assigning participants to conditions in the same way, etc.). While the Generalized Linear Mixed Model I used included individual variation as a factor, it did not take into account the different set of stimuli used in Experiment 2, and thus the results it yields should be interpreted with caution.

A third experiment found no evidence of learning in an *aXb* language where the *a* element was Emphasized and the *b* element Functional, consistent with the hypothesis that NADs are learned not just based on the perceptual/prosodic distinctiveness of the dependent elements, but also on the similarity between the dependent elements. Further research is required to put this finding in context and show whether the similarity between the dependent elements is essential to the mechanism that facilitates the dependency between them.

2.9.1 NAD-learning Guided by Gestalt Principles of Perception

If NADs are acquired based on a mechanism that simply computes co-occurrence probabilities between non-adjacent units, why would this mechanism be affected by the acoustic/prosodic properties of these units? The results pattern with the findings of Newport & Aslin (2004), Creel, Newport & Aslin (2004) and Gebhart, Newport & Aslin (2009), who propose that NAD-learning is facilitated in contexts where Gestalt principles of perception allow the dependent elements to be somehow grouped together, on a separate representational level, based on their perceptual distinctiveness from the intervening material. Thus, participants in the Emphasized Condition and the Functional Condition could have detected the dependencies due to an initial bias to group them together – this bias was due both to the difference in prosodic properties between the target elements and the intervening material, and to the perceptual similarity between the target *a* and *b* elements. Hence, the dependent elements in the Lexical Condition were not distinctive enough to be grouped together, as *a*, *b* and *X* elements all had the same prosodic status.

In addition to the finding that dependency-detection is facilitated by Gestalt principles of perception, I also found evidence that the domain where these Gestalt principles apply is the domain of prosody. Two important aspects of the findings support the conclusion that it is at the level of prosody that elements are grouped together based on similarity. Firstly, the finding that in Experiment 2, when the segmentation confound was eliminated, a significant interaction was found between Pause Version and Acoustic Condition, suggesting that it was not only the acoustic properties of the dependent elements that modulated learning, but also the way in which these elements integrated into the prosodic contour of the strings. Participants performed better where the inter-word pauses were matched to the prosodic status of the dependent elements (short, 100 ms pauses for functional-like nonce words – prosodic clitics – and longer, 250 ms pauses for prosodically marked –emphasized – nonce words), suggesting that properly marking the prosodic status of elements was crucial to the learning mechanism.

Secondly, participants' success in acquiring dependencies in the New Functional Condition with 100 ms pauses is interesting in itself. Although in Experiment 2 the *a/b* elements were recorded so as to integrate into a natural phrasal prosodic contour, because they were no longer recorded in precisely the same slot in their carrier sentences, their acoustic properties now differed. While *a* elements had a rising pitch contour, *b* elements had a declining one; furthermore, because *b* elements were now phrase-final, they were subject to final lengthening, meaning the duration of *b* elements was longer than that of *a* elements, as can be observed in Table A2.3 of the Appendix. If *a_b* dependencies in this experiment were detected based on the acoustic similarity of *a/b*, then the difference in duration and pitch contour between *a/b* in the New Functional 100 ms Condition should not have facilitated the detection of the dependencies. Instead, learners seem to have abstracted away from the acoustic differences between the *a* and *b* classes and categorized them both as prosodic clitics, with different positions in a larger (phrasal) prosodic unit.

Another proposal from Newport & Aslin (2004) that seems to be borne out by the current findings is the fact that NAD-learning is constrained in ways compatible with the properties of natural languages. In the present experiments, one of the two Conditions where participants showed above-chance performance was the one that emulated the prosodic properties of morpho-syntactic dependencies ubiquitous in many natural languages: the dependent elements were perceptually similar to functors, and were separated by minimal (100 ms) pauses from the intervening lexical-sounding nonce words, in strings that aimed to resemble the rising-falling pitch contour of a natural phrasal unit. The fact that this Condition enabled learning, when the Lexical Condition showed no evidence of doing so, suggests that the prosody and perceptual cues that natural languages readily exhibit may be well-suited for the detection of co-occurrence patterns between functional morphemes.

Note however, that while both the Emphasized 250 ms and the New Functional 100 ms Conditions produced above-chance accuracy scores, the effect sizes reported in Table C2.1 show a somewhat smaller effect in the New Functional 100 ms Condition. It may be, therefore, that although prominence of the dependent elements is not a crucial factor to NAD-learning, it does facilitate the detection of dependencies. The potential importance of acoustic prominence as a perceptual cue to NADs may indicate that the properties of function words in natural languages do not, perhaps, make them ideal for learning dependencies between them. However, it does not detract from the core finding that non-adjacent dependencies are learned even when dependent elements are perceptually less salient than the intervening material, and that dependencies emulating the properties of natural-language morpho-syntactic dependencies are learned reliably in an artificial grammar learning setting.

2.9.2 Caveats and Questions for Future Research

Several important points should be mentioned in order to nuance our understanding of the findings in this study, and to inform future research on the topic. First and foremost, while I found an overall interaction between Acoustic and Pause condition, suggesting an effect of prosody and superior learning in conditions with natural prosodic contours and perceptually distinctive *a_b*, this does not mean learning in the other conditions is not possible (merely that it is significantly more difficult). I found that prosodic cues *facilitate* the detection of non-adjacent regularities, but failure to detect a learning effect in the remaining conditions could be due simply to a lack of power. With larger sample sizes it could be possible to find a reliable but small learning effect even in the Lexical Condition, or with the Emphasized *a* – Lexical *X* – Functional *b* Condition employed in Experiment 3. The purpose of the study, however, was to find a reliable *difference* between the conditions, indicative of a facilitative role for specific prosodic cues. A potential replication of this study with a larger sample size should find the same Acoustic by Pause condition interaction, with better learning in the conditions where the dependent elements were perceptually distinct compared to the intervening material, and integrated into a naturalistic prosodic contour.

Secondly, I attributed the above-chance performances found in the Emphasized and Functional Conditions to Gestalt principles of perception that rely on the prosodic contrast between the target elements and the intervening material. However, an alternative and perhaps more parsimonious explanation for the current findings could be that learning was promoted by the existence of *any* prosodic contrast. Specifically, in the Lexical Condition all the words in the *aXb* strings were recorded in exactly the same context, and therefore had the same prosody. The strings were therefore monotonous, and the absence of prosodic contrast could have rendered not just the *a_b* dependency less salient, but could have rendered the *aXb* strings in their entirety less salient. Thus, participants in the Emphasized 250 ms and Functional 100 ms Conditions might have *attended* to the stimuli more readily than participants in the Lexical Conditions, because the former presented them with stimuli with a more salient prosodic contour, or more ‘natural’-sounding. Since there is no independent measure to ascertain attention during the familiarization phase, we cannot immediately dismiss the possibility that participants were simply more engaged in listening to the languages that sounded closer to what a natural language might sound like (i.e. not monotonous, but with a natural prosodic contour). However, if participants were paying more attention in conditions where the *aXb* strings were overall more prosodically salient, it is not clear why participants in Experiment 3 wouldn't also exhibit above-chance learning, since this condition too strived to emulate natural language prosody.

Finally, it is important to point out once again that participants in these experiments were adults, who might have treated the input as a natural language, and may have drawn from their already established experience of natural language(s) the notion that perceptually distinct/reduced elements are generally likely to entertain (morpho-syntactic) dependencies. The stimuli in these experiments were Dutch-sounding nonce words, recorded by a Dutch native speaker; the New Functional Condition with 100 ms pauses was meant to best emulate morpho-syntactic dependencies in Dutch, which was the native language of the participants. It is, therefore, not far-fetched to assume that participants might have been detecting dependencies in the artificial language they heard more easily based on their (perceptual) similarity to dependencies in their own native language.

If adults' detection of the dependencies in the Functional Condition was facilitated by their experience of natural languages exhibiting similar kinds of dependencies, then this may represent a confound in the present study, and obscure the specific way that the NAD-learning mechanism functions in the absence of other biases. One way to eliminate this bias would be to test learning of dependencies between perceptually reduced elements in a non-linguistic domain, visual or auditory. Creel, Newport & Aslin (2004), as well as Gebhart, Newport & Aslin (2009) showed that NAD-learning can be extended to patterns between non-linguistic elements, and that NAD-learning in the non-linguistic domain may also be driven by Gestalt principles of perception. Participants' sensitivity to non-adjacent patterns between perceptually reduced elements in a string of tones or noises would prompt the conclusion that the results of this study may not be entirely determined by linguistic experience, but may reflect a more general property of the NAD-learning mechanism.

Alternatively, the workings of the NAD-learning mechanism may be better evidenced in a study with participants who are less biased by their knowledge of natural language dependencies: infants. In order to establish whether participants in this experiment were relying on their knowledge of natural languages, or purely on the perceptual cues under scrutiny, this study should be replicated with infants, who are capable of dependency-learning but do not yet show knowledge of dependencies in their own language. Note that although behavioral evidence suggests that the NAD-learning mechanism emerges around 15 months (Gómez

& Maye, 2005), neurophysiological evidence suggests that sensitivity to non-adjacent patterns arises as early as 4 months (Friederici, Mueller & Oberecker, 2011). Friederici and colleagues exposed German 4-month-old infants to Italian sentences containing two morpho-syntactic dependencies (*La sorella sta cantando*, 'The sister is singing' vs. *Il fratello puo cantare*, 'The brother can sing') and measured ERP response to sentences containing grammatical vs. ungrammatical dependencies (*La sorella sta cantando* vs. **La sorella sta cantare*). Infants showed a significant positivity 640-1040ms after the onset of the suffix, when the suffix was mismatched with the auxiliary compared to when the two were well-matched. This positivity increased across learning blocks, suggesting that infants were gradually developing sensitivity to the morpho-syntactic dependency, in a language they had never heard before.

Therefore it is also not clear at what age the child should be considered to have no relevant knowledge of NADs in the native language. However, indirect evidence can be obtained by comparing NAD-learning at different ages. If younger infants around the age of 15 months learn NADs only in the Emphasized, and not the Functional Condition, whereas older infants (e.g. 18-24 months) can learn dependencies in both these types of Conditions, it could be shown that the ability to detect NADs when the dependent elements are perceptually non-salient is not an intrinsic property of NAD-learning, but is an acquired ability. Infant research is required to disentangle potential factors contributing to the adult results reported.

Chapter 3

Incremental Learning of Non-Adjacent Dependencies

In the previous chapter I showed that the prosodic properties of functional elements in (at least certain) natural languages could well promote the detection of dependencies between them. Natural languages, therefore, might be optimally designed for the learner to detect the relevant patterns in the speech stream. In this chapter I direct my attention to the question of whether the specific timeline of language acquisition may also be optimally designed to make efficient use of the NAD-learning mechanism. Specifically, I rely on two observations: (i) infants show behavioral evidence of (some) ability to identify functional words in their language around the age of 8-10 months (Shi, Cutler, Werker & Cruickshank, 2006; Shi & Lepage, 2008; Hallé, Durand & de Boysson-Bardies, 2008; Höhle & Weissenborn, 2003, etc.) and to segment functional inflection around the age of 11-16 months (Soderstrom et al., 2007; Marquis & Shi, 2012; Mintz, 2013), and (ii) behavioral studies suggest that awareness of morpho-syntactic dependencies in one's native language develops around the age of 18 months (Santelmann & Jusczyk, 1998; Tincoff, Santelmann & Jusczyk, 2000; Wilsenach & Wijnen, 2004; Höhle et al., 2006; Van Heugten & Johnson, 2010; Van Heugten & Shi, 2010). Taken together these studies suggest that infants may first identify the building blocks (individual morphemes) of morpho-syntactic dependencies, and only subsequently detect the dependencies themselves. If this timeline is accurate, and if the NAD-learning mechanism investigated so far indeed supports the acquisition of morpho-syntactic dependencies, then this makes certain predictions about the workings of the NAD-learning mechanism. The expectation is that NAD-learning could be construed as part of a two-stage, incremental process: an initial stage in which individual elements are identified and encoded separately, and a subsequent (NAD-learning) one in which the co-occurrence patterns between these familiar elements are detected.

In this chapter I explore precisely this prediction. In section 3.1 I introduce the "Less is More", or "Starting Small" hypothesis (Newport, 1990; Elman, 1993), which proposes that some of the structural complexity of natural languages can be learned more efficiently by breaking up the input into smaller chunks. This hypothesis predicts that identifying individual units and detecting local dependencies first is beneficial, or even crucial to the subsequent acquisition of more complex, long-distance patterns. "Starting Small" and its bottom-up approach to learning is therefore consistent with an incremental NAD-learning process that builds representations of remote dependencies of pre-existing representations of the individual dependent elements. Section 3.2, on the other hand, presents an opposing view: the “Starting Big” hypothesis proposes that some types of learning may benefit more from a top-down approach, in which larger chunks of input are first internalized, followed by subsequent segmentation and detection of co-occurrence patterns within those chunks. Starting Big predicts that associative learning mechanisms (such as dependency-learning) are ‘blocked’ by prior familiarity with the individual elements of the dependency. This hypothesis is consistent with a construal of NAD-learning as a single-stage process in which the representations of the dependency and those of the individual dependent elements must be formed simultaneously. Section 3.3 expands on the acquisition timeline for morpho-syntactic dependencies in natural languages, and discusses the relevance of Starting Big or Small for the acquisition of non-adjacent dependencies. Experiments 1-4 investigate the role of “Starting Small” in non-adjacent dependency learning, and the final section offers an overview of the experiments and a discussion of the results.

3.1 The ‘Starting Small’ Hypothesis

Newport (1988) reviewed work on the acquisition of American Sign Language (ASL) by young children and adults, particularly of morphologically complex motion verbs, comprising separate morphemes for the object, path and manner of movement (among others). She observed that the patterns of errors young learners produce are qualitatively different from those of adults. While young children omit specific morphemes, or produce morphemes in isolation, adult learners produce ‘frozen’ forms (unanalyzed morpheme combinations that are highly frequent in the input but inappropriate in the context they are used in), and employ individual morphemes incorrectly and inconsistently. Thus, while children show correct identification of the different morphological units of the complex verbs, adults seem to internalize complex verbs holistically, betraying a lack of sensitivity to their morphological structure. Why would young learners show an advantage over older ones in the morphological analysis of complex verbs, when it is the latter group who are expected to have superior cognitive abilities?

Newport proposed that the cognitive limitations of early age, in particular limitations on perception and memory, storing and processing complex forms, could actually be an advantage in the process of language acquisition. Because young children would find complex verbal expressions difficult to store and process holistically (more so than adults, who have superior working/short-term memory capacities), they must break these forms down into parts and learn these parts individually. This pressure may be what makes early learners more analytical, and could be at the root of the highly systematic representations that are proper to native signers. Newport (1990) elaborated on this idea of *maturational constraints*: she pointed out that both the data cited above and data from second language acquisition suggests that age of acquisition (more than other factors like formal instruction, length of experience, motivation, etc.) is the main determinant of proficiency in a language. This is consistent with Newport's "Less Is More" hypothesis in that only age, and not instruction, experience or learning motivation, is correlated with short-term memory (STM) abilities, and it is also uniquely correlated with the ability to acquire a language. Furthermore, the "Less Is More" approach rests on the idea that the computational effort required to break down complex morphology into its simplex components is much higher when the input is analyzed as a whole, than when the complex units are piecemealed randomly (Goldowsky & Newport, 1990).

Previous research has shown that STM develops gradually throughout childhood. STM capacities in the visual domain have been shown to go through significant development as early as the first year of life (Pelphrey et al., 2004; Rose, Feldman & Jankowski, 2001), a process that has been linked to development in the prefrontal cortex (Nelson, 1995). Verbal STM capacities, as measured through digit/word/non-word recall tasks (Alloway, Gathercole & Pickering, 2006), also show a linear development starting from the age of 4 to 15 (Gathercole, Pickering, Ambridge & Wearing, 2004). Some authors have claimed that it is not the *memory span* that develops from infancy to adulthood, but rather the speed and efficiency with which items are *encoded* (Case, Kurland & Goldberg, 1982), or the emergence of mnemonic or 'chunking' strategies to facilitate memorization (Dempster, 1978). Thus, for instance, if more resources are spent on *encoding* items, fewer resources will be left available for storing multiple items together. Whatever the underlying cause of children's limited memory span may be, the outcome seems to be consistent with Newport's claim: the ability to hold multiple units or larger chunks in memory develops gradually from childhood to adulthood.

Firstly, it is important to note, however, that children's verbal STM is measured using word- or digit-span tests that can only be administered to children around 4 years or older. Therefore, we have a less clear understanding of the development of verbal STM capacities in young infants at the age when they develop sensitivity to the morphology of their native language (1-2 years, see Chapters and 2). Secondly, we must also bear in mind that memory span tests generally assess the ability of a person to retain *and* reproduce a list of familiar items (words or numbers) in the correct order. Such tasks hardly approximate the difference between internalizing morphologically simple vs. morphologically complex linguistic units in language acquisition. In short, it is unclear whether the development of STM abilities in childhood supports the claim that children at the earliest stages of language acquisition are less capable of internalizing morphologically complex chunks and therefore break the input down into smaller, morphologically simplex units.

Elman (1993) extended Newport's "Less Is More" hypothesis (also titled "Starting Small") in a study where he trained a Simple Recurrent Network (SRN) on an artificial language which featured such grammar rules as subject-verb agreement, different verbal argument structures and multiply embedded (subject/object) relative clauses. Learning was significantly improved by providing 'staged' input, in which the network was first familiarized with the simpler sentences and only subsequently with the more complex ones. Furthermore, when the 'working memory' of the network itself was limited and then increased gradually, by controlling the number of previous words that were available when processing the current word, the network also reached a very good level of performance, on trained as well as novel input. Elman explained his results by pointing out that the multiple grammatical rules in the input interacted in complex ways, yielding long-distance dependencies and complex patterns that were too difficult for the network, as they would be for a human learner; yet when the input was graded in terms of complexity, or when its working memory was reduced, it would focus on simple sentences first, where simple rules like agreement or argument structure could be more easily learned. When it finally moved on to more complex input, these rules would already be established and able to explain away some of the complexity of the longer sentences, allowing the network/learner to infer the rest of the grammar rules more easily.[7]

Various studies have tried to evaluate Newport's (1990) and Elman's (1993) claim that reduced memory capacities could actually provide an advantage in learning structures and regularities from the input. Using adult participants, these studies simulated limitations in short-term or working memory in a variety of ways: by distracting participants with a simultaneous (secondary) task that burdened processing capacity; by presenting participants with input of increasing complexity to simulate a gradually expanding STM span; or by correlating participants' actual STM spans with their performance in the learning task. If participants acquired the rules of the input better when their processing/STM

[7] See, however, Rohde & Plaut (1999) who failed to replicate Elman's findings with a similar grammar but different training methods and found the opposite pattern (better learning without staged input) when they introduced semantic constraints within the argument structures of the verbs.

resources were in fact more limited, this would constitute evidence in favor of the "Starting Small" hypothesis.

Cochran, McDonald & Parault (1999) employed American Sign Language to show that limiting participants' attention by a simultaneous secondary task, or exposing them to isolated verbal morphemes, can enhance adult learners' ability to acquire verbal subject/object agreement. Subjects were presented with videos of SVO sentences signed in ASL, and were subsequently tested on their ability to sign either familiar sentences, or new sentences composed of morphemes they had seen in different contexts. Subjects in the 'load' condition had the simultaneous task of counting high-pitched tones played during training. Although in the early stages of learning 'load' learners seemed to perform worse than 'no-load' learners, often omitting morphemes, they subsequently recovered by showing a superior ability to generalize the morphological rules they had acquired to novel contexts, and not internalize fixed forms holistically. Furthermore, learners who practiced signing individual morphemes instead of whole expressions at familiarization also showed superior abilities to employ morphemes in novel contexts. The authors concluded that, in both cases, learners who were attending to smaller units of input were, although at an initial disadvantage, eventually better equipped for capturing the systematicity of the morphology.

Ludden & Gupta (2000) investigated the effect of a simultaneous drawing task on learning an artificial language. Participants who did not have the additional drawing task performed better both in an experiment where they had to segment words from a continuous string based on transitional probabilities between syllables (cf. Saffran et al., 1996), and in an experiment where they were exposed with two-nonce-word pairs and had to detect the 'agreement' within the pairs reflected in the dependency between the 'suffixes' (final syllables) of the two words. The authors concluded that their findings flew in the face of the "Starting Small" hypothesis, as participants whose memory capacities were limited by a simultaneous task did not learn regularities better, or even as well as, participants who were undistracted during familiarization.

Conway, Ellefson & Christiansen (2003) tested the effect of staged input in artificial grammar learning of center-embedded and right-branching recursive structures. They found an advantage to being exposed to input of increasing complexity only with visually-presented, but not auditory stimuli, and concluded that the benefits of starting small applied only specifically to visually-presented, staged input of a relatively high degree of complexity.

Kersten & Earles (2001) showed an advantage for staged familiarization in the visual, word-learning domain. They presented adults with scenes where a mobile entity moved relative to a stationary entity, in different directions and manners. These scenes were depicted by three-word nonce sentences specifying the object, path and manner of movement. Each of the three classes of words (denoting object, path, manner) had a suffix marker (e.g. the path words were *neematig, farnitig, quopatig, segotig, vustutig* and *doochatig*). Participants in the group that was initially presented with one-word, then two-word, and lastly three-word utterances accompanying the visual display learned the mappings better than participants who were exposed to the complete three-word utterances from the very beginning, even though the latter arguably received more (complete) information.

Finally, Kareev, Lieberman & Lev (1997) showed that STM capacity, as measured by a standard digit-span test, correlated negatively with the degree to which rules were over-generalized and used to make predictions. Participants drew red or green envelopes containing coins marked either with an X or an O, and were meant to predict the marking on the coin before opening the envelope. The correlation between the color of the envelope and the mark on the coin was varied. Results showed that participants with lower STM scores used envelope color as a predictor more often compared to participants with higher STM, suggesting that the former perceived the correlations between the color of the envelope and the mark on the coin as stronger. The authors concluded that a lower STM enables a more reliable identification of contingencies in the input, and predicted that learners with STM limitations may be better at filtering out noise in the input and making predictions based on the regularities they have detected.

The studies cited above investigate Newport's (1990) claim in a variety of different ways, and the differences in their methodologies are highly relevant in assessing the implications of their findings. For instance, although Cochran et al. (1999) and Ludden & Gupta (2000) both employ simultaneous tasks as a means to limit learners' processing resources, these tasks are quite different (tone-counting vs. drawing), and thus limit not just the STM span of the learners but also their attention, and the degree to which learning can be explicit. Both tone-counting (Hazeltine et al., 1997) and drawing (Saffran et al., 1997) have been used as tasks that promote incidental or implicit learning. Yet Cochran et al. and Ludden & Gupta do not justify using tasks that purportedly limit (explicit) attention in order to simulate limitations in memory span. More caution is needed in defining the notions we are working with: 'processing resources' can become an umbrella term for a variety of constructs (short-term- or working memory, attention, etc.), and it becomes unclear which of these components, when burdened, do indeed yield superior learning.

Not only do we need to be precise about the manipulations employed to limit processing resources in these studies, but we should also identify the exact type of learning that these manipulations promote or inhibit. In some studies, like

Cochran et al. (1999) and Kersten & Earles (2001), participants are being tested on their ability to break down morphologically complex structures into their simplex units, and acquire stable representations of these units in terms of both form and meaning (e.g. 'the suffix *–ig* attaches to words that denote path'). In other studies like Ludden & Gupta (2000) or Conway et al. (2003), participants were tested on their ability to detect *patterns* between individual units – a fundamentally different process, which is not only based on individuating elements and identifying their correct function, but identifying the rules that connect them. Crucially, these latter studies have conflicting findings, and it is not intuitively straightforward how STM-limitations could stimulate certain types of rule-learning. Finally, Kareev et al. (1997) point to a third aspect of learning that may be (positively) influenced by memory limitations: the ability to deal with *noise* in the input. The claim is that learners faced with limitations in memory and processing resources will simply ignore input strings that are 'noisy', i.e. inconsistent with regularities identified in the previous input and only internalize the input that is more predictable, therefore easier to process. These kinds of learners will perceive statistical regularities as stronger than they actually are, will show robust learning in the face of noise, and will be prone to overgeneralizations, in the same way that children are in the early stages of language acquisition. In recent studies, Newport and colleagues have confirmed that children, unlike adults, are indeed more prone to overgeneralization and ignoring the noise in the input (Hudson Kam & Newport, 2005, 2009; Austin & Newport, 2011).

To conclude this section, Newport and Elman's "Less Is More"/ "Starting Small" hypothesis proposes an interesting link between the sensitive period of language learning and the limited cognitive capacities of early childhood, but it also leaves a series of important questions open. Firstly, it is unclear what the precise components of cognition are whose gradual development supports language acquisition: short-term or working memory, attention, explicit learning, etc. Secondly, these capacities may be difficult to assess directly in the early stages of development (e.g. verbal STM is difficult to measure in pre-verbal infants), and their relationship to language development is by no means clear. Not only do we not know what the STM span is for a 1-year old who begins to say her first words, it is also not clear that this limited STM span specifically limits the *number of morphemes* than can be processed at one time, as Newport would seem to suggest. There is no direct evidence provided by any of the studies summarized above to support the claim that memory limitations actually prompt identification of single-morpheme units from the input, or prevent the internalization of larger chunks – especially since we cannot assess what represents 'too big' a chunk for an infant.

Finally, language acquisition arguably consists of a variety of distinct processes, some of them involving segmentation and identification of the form and function of individual units, others requiring identification of (local or long-distance) patterns and regularities *between* units. It is important to consider that while some of these processes may indeed benefit from a pressure to break the input down into smaller units, others may in fact rely on more holistic representations.

3.2 Reversing the Argument: The "Starting Big" Hypothesis

The "Starting Small" hypothesis seems to view the early learner as an analytical learner, who proceeds bottom-up by first detecting and memorizing the individual units of language and subsequently integrating them into gradually larger and more complex structures. This perspective, however, has been challenged by authors who believe that even very young infants can and do internalize larger chunks of input.

Peters (1983) signaled the heavy presence of fixed/formulaic phrases (e.g. 'That's mine!', 'I don't wanna') in the early speech of children, and proposed that children memorize chunks of various lengths from their initial input, and only subsequently apply further segmentation to reduce these chunks to their simplest units. Peters saw the process of language acquisition as a constant trade-off between Analytic (segmentation) and Gestalt (holistic representation) strategies, and emphasized the large diversity among children in terms of how heavily they rely on one strategy or the other. Usage-based theories of language acquisition have built on this premise, proposing that children start out with holistic, item-based representations, and only subsequently develop abstract grammatical knowledge when they segment and analyze the chunks they have initially stored as units (Tomasello, 2003; Goldberg, 2006).

In her 2009 dissertation, Arnon set out to show not only that young children *do* in fact internalize larger chunks from the input, but also that "Starting Big" rather than "Starting Small" may be the optimal strategy for native language learning. In an experiment testing irregular plural morphology (*tooth – teeth*), she showed that 4-6 year-old children produced the correct irregular form more often when it is embedded in a familiar sentence frame (*Brush your* -) than in isolation, as a response to a question (*What are those?*). She followed Goldberg's (2006) constructionist approach in claiming that children acquire unanalyzed chunks which they then compare to each other so as to identify smaller units (*Brush your teeth, Brush your hair -> teeth, hair*) and derive generalizations (*Brush your* N). She proposed, contrary to Newport (1990), that while native learners adopt a top-down approach, it is L2 learners that build language bottom-up, forming unusual collocations and 'treat[ing] language as less formulaic than it actually is' (Arnon, 2009: 3).

Arnon (2009) also showed that learning certain aspects of morpho-syntax is facilitated by the initial presence of larger chunks. She created a referential artificial language with two noun classes, each exclusively associated with one of two determiners. Noun classes and determiner-noun agreement are notoriously difficult to acquire in the absence of additional cues marking class membership (Braine et al., 1990; Williams & Lovatt, 2003). However, most studies that identified this difficulty familiarized the participants with the individual nouns and determiners prior to exposing them to the nominal phrases. Arnon compared this practice with the effect of familiarizing participants with complex phrases from the very beginning. English-speaking participants in one group were first presented with the noun-determiner pairings in larger carrier sentences, and subsequently familiarized with the nouns in isolation. Participants in a different group were first familiarized with the individual nouns, and only subsequently heard the carrier sentences with the determiner-noun pairings. All participants received exactly the same input, only in a different order. All spoken strings were presented simultaneously with images depicting a person pointing at the object denoted by the nonce noun. Following a distractor-block introduced to avoid recency effects, participants were tested on their knowledge of determiner-noun pairings, as well as of individual nouns. Participants who had heard larger sequences first outperformed their noun-first peers both in identifying the referent of the nouns, and the correct determiner-noun pairs.

Arnon & Ramscar (2012) accounted for this finding by considering the determiner's role as a predictive cue to the upcoming noun: in the noun-first condition, participants learned the association between the noun and the visual object, making the noun perfectly predictable based on the visual cue alone. On the other hand, in the sequence-first condition, participants initially learned to associate the nouns both with determiners and with the visual cue, such that both were employed as predictive cues to the upcoming noun. Arnon and Ramscar claimed that the absence of the determiner in the initial learning phase in the noun-first condition *blocked* the subsequent association between determiner and noun: by the time the learners heard the determiner + noun pairings they had already associated the noun with its referent and the determiner was superfluous as a predictor for the noun.

Arnon & Ramscar's account makes the general prediction that in order to form associations between two elements, A and B, both need to be present simultaneously in the initial input, so that one (A) can be construed as a reliable cue predicting the second (B). If learners become familiarized with A and B independently, then the A-B association will be blocked as neither will be construed as a reliable cue for predicting the occurrence of the other. Applied to language acquisition, this hypothesis predicts that infant learners have to pay attention to larger chunks of input from the very beginning, in order to successfully learn relationships of syntactic dependency between words.

An important distinction needs to be made between the two lines of argumentation that Arnon engages in. One is the question of whether children *do* initially retain larger or smaller chunks from the input, and the other is whether they *should* retain larger or smaller chunks in order to better identify the morphological rules. The first is a difficult point to argue: 4-6 year-old children may store larger (phrasal) chunks in memory, but that does not mean that they do not also store the individual units of these chunks separately, nor does it mean that pre-verbal infants also start out with such chunk-based representations. It is not within the scope of this dissertation to attempt to answer such a difficult question, especially since only indirect evidence can be used to infer how infants represent the input at the moment they set off to uncover its rules and regularities.

The second question, however, seems more tractable: identifying whether different types of learning benefit from Starting Big or Small is testable, and can give us an understanding of which strategy might be more beneficial to language acquisition. In section 3.1, for instance, we saw mixed evidence as to whether limiting processing resources or providing staged input improved the detection of agreement patterns or recursive structures (Elman, 1993; Ludden & Gupta, 2000; Conway et al., 2003). In contrast, in this section I reviewed evidence that presenting complex-to-simple rather than simple-to-complex staged input is more beneficial to the detection of combinatorial properties of determiners and nouns (Arnon, 2009; Arnon & Ramscar, 2012). While "Starting Small" has been claimed to facilitate the identification of individual units in complex structures, "Starting Big" has been shown to benefit the detection of regularities (co-occurrence patterns) between these units.

3.3 Starting Small and Learning Dependencies

In this section I look at morpho-syntactic dependencies, and whether their acquisition proceeds in a top-down or bottom-up manner. As pointed out in Chapter 1, morpho-syntactic dependencies are often instantiated between functional morphemes, a class of elements that emerge relatively late in child utterances but have been shown to play a much more substantial role in early language comprehension (Gerken, Landau & Remez, 1990). In other words, infants are aware of these functional morphemes long before they use them in production. Studies have shown that infants beginning with the age of 8-11 months are sensitive to some of the more frequent functors in their language (Shi & Lepage, 2008), identifying them as units in spoken input, discriminating them from nonsense functors (Shafer, Shucard, Shucard & Gerken, 1998; Hallé, Durand & de Boysson-Bardies, 2008), and using them to segment and learn subsequent material (Höhle & Weissenborn, 2003; Shi, Cutler, Werker & Cruickshank, 2006). Bound functional morphemes, such as verbal or nominal inflection are also acquired around the age of 11-16 months (Soderstrom et al., 2007; Marquis & Shi, 2012; Mintz, 2013). Frequency of occurrence seems to be the main factor modulating the acquisition of functors (Shi & Lepage, 2008; Marquis & Shi, 2012): the more often a functor is heard in the input, the earlier it is likely to be acquired.

Around the age of 14-16 months, children have already refined their knowledge of functors by showing sensitivity to the category-selectional properties of various functors: French and German infants learn that determiners predict the occurrence of a subsequent noun (Höhle, Weissenborn, Kiefer, Schultz & Smitz, 2004; Shi & Melançon, 2010). Therefore, by the age that they first exhibit sensitivity to morpho-syntactic dependencies in their language, around the age of 18-20 months (Santelmann & Jusczyk, 1998; Tincoff, Santelmann & Jusczyk, 2000; Wilsenach & Wijnen, 2004; Höhle et al., 2006; Van Heugten & Johnson, 2010; Van Heugten & Shi, 2010), children are already familiar with at least the forms and combinatorial properties of (some) functors in their native languages.

One question is why sensitivity to morpho-syntactic dependencies arises at the age of one and a half years, when sensitivity to individual functors is already evident within the first year of life. On the one hand, it is possible that while sensitivity to some functors (the most frequent, possibly most perceptually salient ones) indeed arises around 8-11 months, other (bound) functional morphemes, such as, for instance, suffixes, may take longer to acquire (Santelmann, Jusczyk & Huber, 2003). On the other hand, it is possible that the ability to detect dependencies itself is restricted at a younger age: as pointed out in the Introduction, behavioral studies have suggested children before the age of 15 months may have too limited memory capacities to easily track remote dependencies in linguistic input (Santelmann & Juszcyk, 1998; Gómez & Maye, 2005; Lany & Gómez, 2008). However, there is evidence that infants as young as 4 months show sensitivity to morpho-syntactic dependencies, after being exposed to subject-auxiliary-verb combinations in a language they have never heard before. As described in Chapter 1 Friederici, Mueller & Oberecker (2011) showed that very young infants could pick up NADs in a short amount of time (see also

Mueller, Friederici & Männel, 2012), gradually, over several familiarization blocks. They also showed that 4-month olds could learn from naturalistic input, from more complex Subject-auxiliary-verb stem-suffix constructions, and that they seemed to show adult-like sensitivity to the auxiliary-suffix dependencies as a grammatical rule.

Thus, the behavioral studies mentioned in this section seem mainly consistent with a “Starting Small” approach: infants become sensitive to functional elements in the language at a relatively earlier age (~12 months) than they show robust behavioral evidence of a sensitivity to morpho-syntactic dependencies (~18 months). Working or short-term memory capacities have been correlated with NAD-learning: when dependencies are rendered more taxing to working memory they prove more difficult to track (Santelmann & Jusczyk, 1998), whereas when they are rendered less taxing they become learnable to even younger age groups (Lany & Gómez, 2008). However, neuro-cognitive methods show a slightly different picture, as even very young learners with (presumably) very limited memory capacities (4-month-olds) seem to develop sensitivity to NADs after only a few minutes of familiarization. It is as yet unclear whether a sensitive method like EEG could also reveal earlier sensitivity to functional categories in one’s native language. Shafer, Shucard, Shucard and Gerken (1998) showed that 11-month-olds but not 10-month-olds showed different ERP responses to natural texts compared with the same texts where the function words had been replaced with foils.

To conclude, it remains an open question whether children develop a sensitivity to remote dependencies in their language based on pre-existing familiarity with the individual dependent elements, or whether associations between these dependent elements are already being formed at the earliest stage of learning (even while the dependent elements themselves are being identified). While it may be very difficult to establish the simple-to-complex or complex-to-simple progression in natural language acquisition, a more tractable problem is which of these strategies, Starting Big or Small, facilitates the various distributional learning mechanisms that may aid a young learner in discovering the structure of her native language.

3.4 Research Questions

The current study is set up to test the conflicting predictions of the “Starting Small” vs. the “Starting Big” hypotheses with respect to NAD-learning. The main research question is whether staged input is beneficial to NAD-learning: do learners acquire dependencies more easily when they are already familiar with the dependent elements individually? Or is NAD-learning an associative learning mechanism inhibited by ‘blocking’ as a result of prior familiarization with the associated stimuli in isolation? These questions have important implications for how we can view NAD-learning as a tool for language acquisition: if infants start ‘Small’, by identifying individual morphemes and their local combinatorial properties, then their ability to eventually detect more remote dependencies should also benefit from this staged input. If, on the other hand, NAD-learning is inhibited by staged input, then this mechanism can only be a sustainable tool for acquisition as long as acquisition itself starts ‘Big’ – that is, infants can simultaneously acquire the individual units of language *and* the relationships between them.

To test whether NAD-learning is facilitated by “Starting Small” or not, I employ an experimental set-up where participants are initially (pre-)familiarized with short *aY* and *Zb* phrases from which the *a* and *b* elements can easily be extracted. Subsequently, they are exposed to the *aXb* language containing the *a_b* dependencies they are meant to acquire, and then tested on their sensitivity to these dependencies, as before. The transition from smaller to bigger chunks is meant to emulate the acquisition timeline of functional elements: infants seem to first identify the functors and the elements they combine with adjacently, and only subsequently become sensitive to the non-adjacent patterns between these functors.

In Experiment 1 I investigate the role of “Starting Small” by varying the extent to which learners are pre-familiarized with the *a/b* elements, and observing the effect that has on the subsequent detection of *a_b* dependencies. Learners receiving more pre-familiarization should be able to encode the *a/b* elements better. If the “Starting Small” hypothesis is correct, learners that are more familiar with the *a/b* elements should subsequently acquire the *a_b* dependencies more easily. Thus, the learning effect should be directly proportional to the amount of pre-familiarization received. If, on the other hand, the “Starting Big” hypothesis is accurate, the more learners are familiarized with *a/b* elements individually, the stronger the ‘blocking’ effect of pre-familiarization should be. The “Starting Big” hypothesis, therefore, predicts an inverse relationship between learning and amount of pre-familiarization.

Experiment 1, therefore, varies the amount of pre-familiarization to determine its effect on subsequent learning of NADs. Experiment 2 keeps the *length* of pre-familiarization constant but varies its *salience*: if *a/b* elements are more easily learned at familiarization, does this promote or inhibit learning at familiarization? In Experiments 3 and 4 I compare the effect of a pre-familiarization that contains the relevant *a/b* elements to one that does not, and in Experiment 4 I look into potential methodological factors that may explain previous findings.

3.5 Experiment 1 – Varying the Length of Pre-Familiarization

In this experiment I investigate the hypothesis that prior familiarity with the individual elements of a dependency supports or facilitates the acquisition of the dependency. In three separate conditions I varied the degree to which participants are pre-familiarized with the target *a/b* elements of (subsequently presented) *a_b* dependencies, by varying the amount (length) of exposure to these elements in isolation. Participants were pre-familiarized with *aY* and *Zb* strings where *a* and *b* combined freely with various bisyllabic words (*Y*, *Z*, where *Y* and *Z* are different classes each containing 6 bisyllabic nonce-words). A 100 ms pause facilitated the segmentation of *a* or *b* units from the phrases. In this pre-familiarization phase, therefore, participants could learn that *a/b* elements combine productively with bisyllabic words, and that while *a*-elements are phrase-initial, *b*-elements are phrase-final. However, learners at pre-familiarization received no information about a potential relationship between the *a* and *b* elements. After this pre-familiarization phase, learners proceeded to familiarization where they heard longer *aXb* strings and were meant to detect the co-occurrence patterns between *a* and *b*. This was meant to emulate a potential transition that young learners make from focusing on the individual properties of functional morphemes and the elements they combine with adjacently, to detecting more remote, non-adjacent relationships between them. The assumption is made that a longer pre-familiarization phase will create stronger memory traces for the target *a/b* words, and thus show a stronger effect of prior familiarity with *a/b*. Thus, if learning of the *a_b* dependencies improves with longer pre-familiarization, then we can conclude that prior familiarity with the individual dependent elements facilitates dependency-learning. If, on the other hand, learning declines with longer pre-familiarization, then we might conclude that the more familiar the *a/b* elements are to the learner, the less likely it is that the learner will subsequently detect a dependency between them.

3.5.1 Materials

The effect of pre-familiarization was tested in three different Conditions. The first condition had a Short Pre-familiarization phase of 1minute: the 3 *a* tokens were combined with 6 different bisyllabic *Y* nonce words, and the 3 *b* tokens were combined with 6 other *Z* words. There was no overlap between the classes of elements that could combine with *a* and *b* tokens, respectively, and no specific cues marking the distinction between the *Y* and *Z* classes. The 36 *aY/Zb* strings (with 100 ms intra-string pauses) thus obtained were presented once, in an order randomized per participant, with 750ms pauses in between; none of the *Y/Zs* was present in the subsequent familiarization or test phase. The Medium Pre-familiarization condition had a slightly longer pre-familiarization that lasted 2 minutes. This time, all 12 *Y/Zs* were combined exhaustively with both *a* and *b* tokens, yielding 72 pre-familiarization tokens, which were presented once, in the same manner as in the Short Pre-familiarization condition. The Long Pre-familiarization condition had an 8-minute pre-familiarization, which consisted of 4 repetitions of the 72 strings from the Medium Pre-familiarization condition. The familiarization and test phase employed exactly the same materials as in the New Functional Condition with 100 ms pauses (see Chapter 2). The additional 12 *Y/Zs* were taken from Grama et al. (2013), and were recorded in the same conditions as the *Xs* from the New Functional Condition.

3.5.2 Participants

I recruited 75 adult participants (10 male), aged between 18 and 56 (mean age 22.07, SD = 5.83) from the UiL-OTS subject database, via email. Subjects reported having no diagnosed dyslexia, hearing or attention deficits, and were given 5euro each for their participation. One participant was excluded (post hoc) for being familiar with NAD-learning research, and one participant was excluded due to technical problems.

3.5.3 Procedure

The procedure was similar to the one in the New Functional Condition with 100 ms pauses (cf. Chapter 2: participants randomly assign to language L1 or L2 sat in a sound-proof booth coloring a mandala while listening to the language; they were only instructed about the test phase after familiarization was over). However, at the beginning subjects were instructed that there would be a shorter first part, followed by a longer second part. When pre-familiarization was over, subjects had to click on a button on the screen that took them to the familiarization, thus ensuring that they were aware of a separation between the pre-familiarization and the familiarization.

3.5.4 Results

Participant's accuracy in answering the test questions (correct endorsement of strings consistent with the familiarization language and correct rejection of strings inconsistent with familiarization) was used to measure learning of the *aXb* language. Mean accuracy scores for the three Conditions, together with *p*-values of one-sample *t*-tests on these scores are presented in Table C3.1, along with Cohen's *d* computed as within-subject comparison between the endorsement rates for correct ($a_iX'b_i$) and incorrect ($a_iX'b_j$) test strings. A Generalized Linear Mixed Model was run with accuracy (per subject per test item) as (binomial) dependent variable, subjects as random factor and Language (L1, L2) and Condition (Short, Medium and Long-) as fixed factors. There was a significant effect of Condition (F (2,872) = 3.513, p = .030), with post-hoc comparisons yielding a significant difference between the Short and Long Pre-familiarization conditions ($t(2)=-2.532$ p = .009, 95% CI [1.14; 7.85]), and no other significant effects or interactions. Learners in the Short Pre-familiarization condition showed significantly higher accuracy scores than those in the Long Pre-familiarization condition, suggesting longer pre-familiarization had a negative effect on learning.

	Short Pre-familiarization	**Medium Pre-familiarization**	**Long Pre-familiarization**
No. subjects	24	25	24
Mean accuracy	M = 60% (SD = 15.86)	M = 53% (SD = 14.02)	M = 49% (SD = 13.39)
One-sample *t*-test	$t(23) = 3.003$ **p = .006** 95% CI [3; 16]	$t(24) = 1.065$ p = .297 95% CI [-2; 1]	$t(23) = -.508$ p = .616 95% CI [-7; 4]
Endorsement correct	M=77 % (SD= 18.26)	M=67% (SD=21.27)	M=65% (SD=21.03)
Endorsement incorrect	M=58% (SD=27.35)	M=61% (SD=24.4)	M=67% (SD=24.81)
Effect size (Cohen's *d*)	**0.627**	**0.287**	**-0.123**

Table C3.1. Results for the 3 Pre-Familiarization conditions with Mean Accuracy of response, Endorsement for correct vs. incorrect test strings, *p*-values for one-sample *t*-test comparing Mean accuracy to chance and effect size (Cohen's *d* for the within-subject comparison between Endorsement correct vs. incorrect)

3.5.5 Discussion

I tested the effect of prior knowledge of individual dependent elements, on adults' ability to learn the dependencies between those elements. NAD-learning was tested in three different conditions with pre-familiarization phases of increasing lengths but equal complexity: the relevant dependent elements were presented embedded in two-word combinations where the other word was a bisyllabic element selected from a set that did not figure in the subsequent familiarization or test. If prior knowledge of dependent elements boosts dependency-learning, as predicted by the "Starting Small" hypothesis, we should see an increase in performance across the three Conditions: an increasingly longer pre-familiarization phase should facilitate the subsequent detection of patterns between the elements.

Contrary to this prediction, the results show a significant decline in performance across the three conditions: an overall drop in the effect size of learning and a significantly poorer performance with Long compared to Short Pre-familiarization. This suggests that prior familiarization with the dependent elements has interfered with the acquisition of dependencies between them, to the point of making the learning effect disappear. If anything, such a pattern of results is more in line with the "Starting Big" hypothesis, which suggests that being familiarized with individual elements that enter into NADs separately blocks the subsequent detection of any associations between them.

Nevertheless, when interpreting these results it is important to point out that what varied across the three conditions tested here was not just the amount of input that the learners received at pre-familiarization, but also the amount of time learners spent in the testing booth, listening to nonce phrases while performing the coloring task, prior to the familiarization phase. In other words, a longer pre-familiarization phase might have taxed the participants' attentional resources, so that they were less able to attend to the structure of the language presented at familiarization. In order to ensure that it was really familiarity with the dependent elements that inhibited learning at familiarization, Experiment 2 set out to eliminate length of pre-familiarization as a confound.

The logic for Experiment 2 relies on the observation that learning in Condition 1 of Experiment 1 (with only a minute-long pre-familiarization) is still robust, and comparable in effect size to learning from the same stimuli without any pre-familiarization phase (i.e. learning in the New Functional 100 ms Condition from Chapter 2, Experiment 2, which yielded an average accuracy rate of 57.99%, with a Cohen's *d* of .516). This suggests that when the *a/b* elements are functional-sounding pseudo-words, 1 minute of pre-familiarization is not enough to have a negative effect on learning.

In Experiment 2 I also tested the effect of 1 minute of pre-familiarization, but increased the prosodic *salience* of the *a/b* elements. The assumption behind this is that making the *a/b* tokens more perceptually salient would facilitate their encoding and memorization at pre-familiarization. This would, in turn, enhance the effect of pre-familiarization on subsequent learning at familiarization. Thus, if prior familiarity with the *a/b* tokens is detrimental to the subsequent detection of *a_b* dependencies, then making these tokens easier to memorize (by increasing their perceptual salience) should have a negative effect on learning dependencies between them.

In a 2x2 design I compare NAD-learning with and without pre-familiarization in two Acoustic Conditions: one in which the *a/b* elements are not perceptually salient (Functional) and one in which they are salient (Emphasized). If my assumption is wrong and higher perceptual salience does not facilitate learning of *a/b* at pre-familiarization, then a 1-minute pre-familiarization phase should have the same effect irrespective of Acoustic Condition. If, on the other hand, the perceptual salience of *a/b* does facilitate learning at pre-familiarization, then having a pre-familiarization in the Emphasized condition should have a greater effect on learning at familiarization. If *a/b* are more easily learned at pre-familiarization due to their higher perceptual salience, and if this learning is detrimental to learning at familiarization, then we predict the following: while introducing a Short Pre-familiarization phase in the Functional condition has had no discernible effect on learning, introducing the same Pre-familiarization in an Emphasized condition should show a significant decline in learning.

In short, in Experiment 2 instead of varying the *amount* of pre-familiarization, I varied the *salience* of that pre-familiarization, while keeping the length constant, to assess the impact that learning at pre-familiarization has on learning at familiarization.

3.6 Experiment 2 – Emphasized Short Pre-Familiarization

3.6.1 Materials

The stimuli from the Emphasized Condition in Chapter 2 were used. The familiarization and test phase were identical to that condition. In the No Pre-familiarization Condition the Emphasized Condition in Chapter 2 was simply replicated, with exactly the same stimuli and procedure. For the Pre-familiarization Condition, I spliced the emphasized *a* and *b* tokens and combined them with the 6Xs and 6Ys used in Condition 1 Experiment 1, respectively, for a total of 36 *aX/Yb* strings. The No Pre-familiarization and the Pre-familiarization Conditions received an identical set of instructions: to listen to the stimuli passively while coloring a mandala. Participants in the Pre-familiarization Condition were not explicitly instructed about the existence of a pre-familiarization phase, so as to keep the instructions the same between the two Conditions: instead, there was a pause of maximally 20 seconds between pre-familiarization and familiarization, during which the sound stopped and the 'Continue' button appeared on the screen. If participants pressed this button during this pause, they would continue to familiarization. If not, at the end of the pause the experimenter pressed the button and the experiment continued to familiarization.

3.6.2 Participants

I recruited 48 participants (12 male, mean age 20.5, age range 18-51). Four participants were excluded (post hoc) as during the debriefing at the end of the experiment they indicated they were familiar with NAD-learning research. Of the remaining participants, 20 were assigned to the No Pre-familiarization condition and 24 to the Short Pre-familiarization condition.

3.6.3 Results

Accuracy scores, effect sizes and *p*-values for the No Pre-familiarization and Short-Pre-familiarization Conditions are presented in Table C3.2. In the No Pre-familiarization condition, a one-sample Kolmogorov-Smirnov test showed that the accuracy rates were not normally distributed (p=.022) so I ran a one-sample Wilcoxon Signed-Rank test which showed significant above-chance performance (Median=58%, SE=15.839, Z=2.336, p=.019). In the Short Pre-familiarization condition, the accuracy rates were not normally distributed (p <.001): a one-sample Wilcoxon Signed Rank test showed that the accuracy rate was not significantly above 50% chance (Median=50%, SE=19.108, Z=.654, p=.513).

A Generalized Linear Mixed Model with accuracy as dependent variable, subjects as random factor and Pre-familiarization (None vs. Short) and Language (L1 vL2) as fixed factors showed a significant negative effect of Pre-familiarization (F (1,537)= 4.069, p=.044) and no effect of Language (p=.157). I also compared the results of the two conditions tested here with the results of learning with/without pre-familiarization with Functional-sounding *a/b* words, to determine whether the effect of pre-familiarization varied with the salience of *a/b*. A Generalized Linear Mixed Model with accuracy as dependent variable, subjects as random factor and Acoustic Condition (Functional vs. Emphasized), Pre-familiarization (No Pre-familiarization vs. Short Pre-familarization), Language (L1 vs. L2) and Acoustic Condition * Pre-familiarization, as well as Acoustic Condition * Language as fixed factors yielded a near-significant interaction Acoustic Condition * Pre-familiarization ($F(1, 1111)=3.494$, $p=.062$), and no other significant effects.

Pre-familiarization	No Pre-fam.	Short Pre-fam.	No Pre-fam. (Exp. 2, Ch 2)	Short Pre-fam. (Exp 1)
Acoustic Condition	EMPH	EMPH	FUNCT	FUNCT
Participants	20	24	25	24
Mean Accuracy	65% (SD=24.12)	53.12% (SD=17.68)	57.99% (SD=15.63)	59.72% (SD=15.86)
p-value	.019*	.513	.02*	.006*
Effect size (Cohen's *d)*	**.624**	**0.18**	**0.516**	**0.627**

Table C3.2. Results for the Emphasized and Functional Conditions, with or without Short Pre-Familiarization with Mean Accuracy and *p*-values of one-sample *t*-tests comparing Mean accuracy with chance (or nonparametric Wilcoxon Signed-Rank tests where accuracy scores were not normally distributed), and with effect sizes (Cohen's *d*).

3.6.4 Discussion

In this experiment I investigated whether perceptual salience plays a role in the way pre-familiarization affects the learning of NADs. Results showed that a one-minute-long pre-familiarization phase has a powerful negative impact on learning performance when the *a/b* dependent elements are perceptually salient: learning in the Emphasized Condition with a one-minute pre-familiarization was significantly poorer than without any pre-familiarization. On the other hand, previous findings suggest that learning was above chance both with- and without pre-familiarization when the dependent elements *a/b* were functional-sounding, therefore less salient. In other words, salient *a/b* words at pre-familiarization blocked learning at familiarization. In this experiment I assumed that participants exposed to the more salient *a/b* elements learned these words better at pre-familiarization, although there was no direct test to measure memory for *a/b* tokens with Emphasized vs. Functional stimuli. Nevertheless, if the difference in the prosodic features of the *a/b* tokens had not affected learning at pre-familiarization at all, we would not have expected a near-significant interaction effect between Acoustic Condition and Pre-familiarization.

Taken together, Experiments 1 and 2 suggest that any attempt to facilitate learning at pre-familiarization (by increasing either the *length* or the *salience* of pre-familiarization) engenders a decrease in learning at familiarization. These results are inconsistent with the "Starting Small" hypothesis but can be accommodated within the "Starting Big" hypothesis: learners are not aided by prior familiarization with the individual words when learning patterns between them. In fact, being familiarized with the words beforehand inhibits learning the patterns between them.

However, these results are compatible with an alternative explanation as well – namely that the simple *presence* of a learning phase prior to familiarization, whatever the words presented in that phase, may be in itself detrimental to learning. In this experimental set-up, I require learners to first attend to a language of two-word phrases, where participants can learn about individual words, their relative position in a phrase (*a* words are phrase-initial and *b* phrase-final), or the type of words that they combine with. Then I change the input to a grammar where they are meant to shift their attention to non-adjacent regularities. This transition happens within minutes, with virtually no pause in between the two stages of input.

One consequence of this set-up might be that the pre-familiarization phase simply introduces *noise* in the input. Because the pre-familiarization and familiarization follow each other seamlessly, it might be that learners simply cannot shift to processing non-adjacent dependencies (at familiarization) if the pre-familiarization phase has primed them to track adjacent relationships. Instead, they may be conditioned by the pre-familiarization to continue attending to adjacent relationships. Thus, it may be that what we are seeing as a negative pre-familiarization effect (NPE) is not due to prior familiarity with the relevant *a/b* words, but to the difficulty of transitioning from one type of input (with only adjacent relationships and without non-adjacent regularities) to another (with non-adjacent regularities). Note that this does not necessarily invalidate the "Starting Small" hypothesis with respect to natural language acquisition in infants. Because their transition from simple to complex is more gradual, and because the ability to detect NADs itself may develop later than the ability to detect adjacent regularities (whereas adults have that ability at pre-familiarization and may already decide it's not a useful strategy to analyze the input), infants may be more flexible in the learning strategies they apply, while adults may become more easily and quickly entrenched in one learning strategy.

In Experiment 3 I control for this possible confound in my methodology in two ways: firstly, the pre-familiarization phase will present the target *a/b* words in isolation. Simple Pre-familiarization provides no cues as to what regularities might be relevant in the subsequent language. It is merely a list-like enumeration of the target elements that figure in non-adjacent patterns at familiarization. Thus, a simple, structure-less pre-familiarization should not direct participants towards a certain type of relationship (adjacent) and distract them from other (non-adjacent) regularities, but should allow participants to look for both adjacent and non-adjacent regularities in the subsequent, structured input.

Secondly, I disentangle the effect of prior familiarity from the effect of pre-familiarization (simply being exposed to a prior learning phase) per se. While participants in an Experimental Condition will be pre-familiarized for 1 minute with a randomized list of *a/b* words, participants in the Control Condition will receive the same type of pre-familiarization but with a list of 6 monosyllabic words that will not appear subsequently at familiarization or test. If prior familiarity with the dependent words has a negative impact on learning the dependencies, then participants in the Control Condition should show significantly better learning than participants in the Experimental Condition. If the mere presence of a pre-familiarization phase (irrespective of its content) hinders learning at familiarization, then learning should also be negatively affected in the Control Condition, compared to what we might expect if there were no pre-familiarization phase.

3.7 Experiment 3 – Simple Pre-Familiarization

3.7.1 Materials

Stimuli from the New Functional Condition with 100 ms pauses (Chapter 2) were used again. I exposed participants to 12 repetitions of all 6 individual *a/b* tokens from the familiarization phase (3 *a*s and 3 *b*s), separated by 750ms pauses, in an order randomized per participant. The familiarization and test phase were the same as before. In addition, I tested a Control Condition in which 6 monosyllabic nonsense words (*blim, floon, jaaf, niem, snet, troen*) with the same acoustic properties as the functional *a/b* tokens were used at pre-familiarization, while the familiarization and test phase were identical to the Experimental Condition. This was done in order to pit the effect of familiarity with the target individual elements against the effect of having a pre-familiarization phase in itself. If familiarity with dependent elements prior to learning the dependency is detrimental to learning, then we should see a significant difference in learning between the Control and the Experimental Condition: whereas learning should be significant in the Control Condition, it should be inhibited in the Experimental one.

3.7.2 Participants

I recruited 44 participants (10 male; age range 18-31, mean age 20.5) in the same way as before. The procedure was identical to that in Experiment 2. While 20 participants were assigned to the Control Condition, 24 participants were assigned to the Experimental Condition.

3.7.3 Results

In the Control Condition, participants responded correctly to the test items 50.69% of the time (SD=13.44). A one-sample Kolmogorov-Smirnov test showed that the distribution of accuracy scores was not normal (p = .004), so I used a one-sample Wilcoxon Signed-Rank test which showed that the mean accuracy rate was not significantly different from chance (Median=50, SE=12.619, Z = .238, p = .812). In the Experimental Condition, participants' mean accuracy was 50.69% (SD=14.31), a one-sample Kolmogorov-Smirnov test showed that the distribution of accuracy scores was not normal (p =.013), so I again used a one-sample Wilcoxon Signed-Rank test, which showed that the mean accuracy rate was not significantly different from chance (Median=50, SE=22.257, Z = .225, p = .822). A Generalized Linear Mixed Model with accuracy as dependent variable, subjects as random factor and Condition and Language as fixed factors showed no significant effect of Condition (p =.949) and no significant effect of Language (p =.140).

3.7.4 Discussion

This experiment explored the possibility that with simpler and shorter pre-familiarization, learners might not show a disadvantage of being pre-exposed to the tokens that would later enter the dependency rules they had to learn at familiarization. With a simple pre-familiarization consisting of a list-like enumeration of the target *a/b* elements, subsequent learning of *a_b* dependencies was completely at chance. However even when the pre-familiarization consisted of different monosyllabic words that would not subsequently occur at familiarization or test, learning was equally degraded.

The results of this third experiment are puzzling and raise concerns about the validity of Experiments 1 and 2. Why would a one-minute pre-familiarization with a list of completely irrelevant words inhibit the subsequent learning of an *aXb* language? Three logical possibilities arise: (i) that it is the *nature* of the pre-familiarization phase which affects learning at familiarization; (ii) that the very *occurrence* of a pre-familiarization phase is an impediment to learning in the familiarization phase, and finally (iii) that the *transition* between pre-familiarization and familiarization compromised learning. I examine these possibilities in turn below.

The first possibility (i) is that it is not the presence of pre-familiarization itself which degrades learning, but rather the list-wise nature of the Simple Pre-familiarization. This amounts to a reversal of the rationale I provided for Experiment 3: while it is possible that presenting two-word phrases at pre-familiarization tunes the learner's attention to adjacent instead of non-adjacent relationships, it is also possible that a pre-familiarization enumerating isolated words lead the participants to engage more in word-learning than rule-learning. The distinction between *words* and (grammar) *rules* has been intensely studied over the past few decades (Pinker, 1998; Ullman 2001, 2004, among others), and crucial differences have been found between them in how they are learned, processed and represented in the brain. Artificial grammar studies have shown that segmenting words from the input and abstracting rules engage fundamentally different computational mechanisms (Peña, Bonatti, Nespor & Mehler, 2002; Endress & Bonatti, 2006). Item-based and rule-based learning processes also elicit systematic differences in electrophysiological response registered at the scalp (Diego-Balaguer, Toro, Rodriguez-Fornells & Bachoud-Levi, 2007) as well as in the synchronization of neuronal assemblies (Diego-Balaguer, Fuentemilla & Rodriguez-Fornells, 2010). De Diego-Balaguer et al. (2010) showed that in an artificial language experiment which required the detection of both word-units and non-adjacent dependency-rules, the learners who exhibited higher sensitivity to words than rules showed more localized patterns of synchronization, whereas the more rule-sensitive learners showed an increase in synchronization between different (frontal, temporal and parietal) regions in the brain. While poor rule-learners consolidated the memory traces of the words they were segmenting, good rule-learners showed a shift in attentional resources from segmenting words to identifying patterns. Rule-learning and word-identification, therefore, seem to be distinct, yet competing mechanisms, such that when one is activated the other is inhibited, and vice-versa. Thus, it is possible that presenting participants with a list of words to be memorized activates item-based learning mechanisms, but inhibits the subsequent detection of rule-like patterns in the input.

The second possibility I consider (ii) is that learning at familiarization could have deteriorated as a consequence of the very fact that learners spent one minute before it listening to a different type of material. Under this assumption, learning of an artificial language would be optimal only when no other learning phase preceded it. This is less likely given that in Experiment 1 participants who received a 1-minute pre-familiarization with *aX/Yb* strings showed significant learning, comparable to learners in an identical condition that lacked a pre-familiarization phase. The negative pre-familiarization effect, therefore, seems to arise only in particular circumstances that may be related either to the nature of the pre-familiarization itself or the nature of the *transition* (seamless or explicitly cued) between pre-familiarization and familiarization (iii).

To examine this third possibility (the transition between learning phases), it might be useful to observe the analogy between the experiments above and studies looking at participants learning two distinct grammars in quick succession. Gebhart, Aslin & Newport (2009) showed that the consequent presentation of two separate artificial grammars A and B, with different vocabularies and either similar or different structures to be learned, always exhibited a *primacy effect*: language A would be learned as well as if it had been presented in isolation, whereas language B was not learned at all. However, when a sufficiently salient cue (30s pause) was introduced to separate the presentation of the two languages, both A and B were learned equally well. The authors suggest that this primacy effect is determined by the learners' limited ability to detect changes in the input. Namely, the participants tune in to language A, acquire its patterns, and internalize its rules. When the input changes, listeners are committed to the representations of language A and are not easily influenced by the subsequent changes in structure. That is, humans can learn from quite little input, but once they form representations of the structures they are hearing, these representations are quite stable and difficult to alter – a little noise in the signal should not affect them.

However, Gebhart and colleagues showed that if language B exposure is tripled, then learners become able to form a separate, stable representation for the structure of language B as well, without affecting the representations of language A. Zinszer & Weiss (2013) replicated and extended Gebhart et al.'s findings. They found that while presenting single blocks of exposure for languages A and B showed a primacy effect, when they presented alternating blocks of A and B, both languages were learned, even when presented in immediate succession with no pause or transition cue. Zinszer & Weiss showed that contextual cues were not necessary for learners to detect the difference between A and B, as long as multiple transitions between the languages existed, but in their design, languages A and B contained exactly the same type of regularities (segmentation cues based on transitional probabilities between vowels) but instantiated with a different vocabulary.

Mitchell & Weiss (2010) also investigated learners' ability to form two distinct systems based on the properties of the input. They employed continuous streams of syllables which were segmentable into tri-syllabic units based on the transitional probabilities between individual syllables (cf. Saffran et al. 1998). Two different languages, with partially overlapping inventories of syllables were constructed, such as the same syllable could be word-final in one language, but not the other. The languages were presented alternatingly, in such a way that treating the input as one language would yield less reliable segmentation cues than treating it as two separate systems. Mitchell and Weiss found that when the two languages were synchronized with videos of two different speakers pronouncing them, such that the learner could associate each language with a different speaker, both were learned equally well. Yet when this cue was removed, or when a different contextual cue (background color instead of human speakers) was associated with each language, the primacy effect re-appeared.

The studies above show that the serial presentation of two distinct grammars may pose problems for the acquisition of the second grammar in the absence of powerful cues marking the transition or dissociation between the two grammars. In the experiments presented above the methods used to cue the transition between pre-familiarization and familiarization varied. In Experiment 1, for instance, participants were explicitly instructed that learning consisted of two different phases and were prompted to click the 'Continue' button at the end of the first. In Experiments 2 and 3, participants were not explicitly told about the existence of a pre-familiarization phase. Once the pre-familiarization ended and the sounds stopped, they were given a 20-second period to react and click the 'Continue' button, otherwise the button was pressed for them and the experiment continued automatically. Although Gebhart, Aslin & Newport (2009) proposed that even a short pause and a much longer exposure to the second language (in our set-up the familiarization is up to 15 times as long as the pre-familiarization) should provide sufficient cues to cancel out the primacy effect, in their study participants were not given a simultaneous task to complete during learning. In fact, their participants were instructed to listen carefully to the input, and when the pause cue was present they were also specifically told about the existence of two distinct learning phases. In an incidental learning situation such as ours, with low levels of explicit attention to the input, it may be that the transition from one grammar to the other may be even harder to detect.

The literature cited above seems to suggest that learning an artificial grammar is possible even when preceded by another learning phase, as long as sufficient cues are provided as to the separation of the two learning phases. In Experiment 4 I follow suit and compare the effect of pre-familiarization on NAD-learning with two different sets of procedures: the explicit instructions that were given to participants in Experiment 1 underlining the existence of separate learning phases, and the instructions given in Experiments 2 and 3 which did not mention the existence of a separate pre-familiarization phase. I also continue to investigate the effect of prior familiarity with the target words *a/b* by comparing the effect of a pre-familiarization that contains these words with the effect of a pre-familiarization that contains no pseudo-words that will appear later at familiarization or test. A 2x2 design with Instructions and Condition (Experimental, Control) as between-subjects factors should answer both research questions.

If the negative effect of pre-familiarization revealed in Experiment 3 is similar to the primacy effect identified by Gebhart, Aslin & Newport (2009), then I expect a main effect of Instruction: participants who received more explicit instructions about the existence of two separate learning phases should show better learning at familiarization. If prior familiarity with the *a/b* words affects learning of *a_b* dependencies, then we should see significantly better learning in the Control Condition than in the Experimental Condition. If, on the other hand, prior familiarity with *a/b* is beneficial to learning *a_b* dependencies, the opposite pattern should emerge.

3.8 Experiment 4: Controlling for Pre-Familiarization and Instruction

Two different pre-familiarization Conditions (Control and Experimental) were tested with either explicit Instructions or No Instructions as to the existence of two separate learning phases (pre-familiarization and familiarization). I employed a Short (1-minute) Pre-familiarization where participants were exposed to 36 *aY/Zb* strings. In order to maximize the learning effect while avoiding a lengthier pre-familiarization, the stimuli were Emphasized *a/b* tokens.

Thus, the Emphasized Condition with Short Pre-familiarization in Experiment 2 coincided with our Experimental No Instructions condition. I did not replicate this condition but tested the remaining 3 conditions (Control with and without explicit Instructions, Experimental with explicit Instructions), and compared them to the former.

3.8.1 Materials

The materials for the Experimental Condition (with new Instructions) were the same as in Experiment 2. For the Control condition, 6 monosyllabic words (*daap, drig, floon, niem, snet, troen*) were recorded in the same way as the emphasized *a/b* tokens and embedded in *wX/Zv* phrases. The 6 *Y*s and 6*Z*s were identical to the pre-familiarization phase in the Experimental Condition. The Control and Experimental Conditions were thus identical, except that in the Control Condition the target *a/b* tokens at pre-familiarization were replaced with other monosyllabic words, which were not heard again during familiarization or test.

3.8.2 Participants

Eighty-five participants were recruited in the same way as before (28 male; age range 18-54, mean age 23).

3.8.3 Procedure

The procedure in the Control Condition with No Instructions was identical to the procedure in Experiment 2: participants were not told in advance of the existence of two separate learning phases, but at the end of pre-familiarization the sound stopped and a message appeared on the screen that asked them to press a 'Continue' button to move on to the next phase. If participants pressed this button, they moved on directly to familiarization. If they did not press this button within a 20-second period, the experimenter pressed the button for them and the experiment continued automatically to familiarization.

The procedure in the Control and Experimental Conditions with explicit Instructions differed only in that at the start of the experiments the participants were told that there would first be a short learning phase at the end of which there would be a message asking them to press a button to continue to the next part. They were told to press that button and go to the next learning phase, which would be longer. At the end of the second learning phase they had to stop and wait for the experimenter to come in with new instructions. They were told to color the mandala throughout both learning phases.

3.8.4 Results

A Generalized Linear Mixed Model with accuracy as dependent variable, subjects as random factor and Condition (Experimental vs. Control), Instructions (No Instructions vs. Explicit Instructions), Language (L1, L2) and Condition * Instructions interaction yielded a near-significant effect of Instructions ($F(1,1315) = 3.301$, $p=.069$) and no other significant effect or interaction. Results for the 4 cells of the 2x2 design can be seen in Table C3.3.

		No Instructions	**Instructions**
Control (*wY/Zv* at pre-familiarization)	Participants	25	35
	Mean Accuracy	54 % (SD=13.4) $p = .158$	57% (SD=16.8) $p = .021$*
	Effect size (Cohen's *d*)	***d* = .304**	***d* = .410**
Experimental (*aY/Zb* at pre-familiarization)	Participants	24	25
	Mean Accuracy	53.12% (SD=17.68) $p = .513$	62.67% (SD=21.53) $p = .007$**
	Effect size (Cohen's *d*)	***d* = .18**	***d* = .595**

Table C3.3. Results for the 2x2 design (Condition x Instructions) in Experiment 4, with Mean Accuracy and *p*-values of one-sample *t*-tests comparing Mean Accuracy to chance (or nonparametric Wilcoxon Signed-Rank tests where accuracy scores were not normally distributed), and with effect sizes (Cohen's *d*).

3.8.5 Discussion

In this fourth experiment I set out to test both the role of prior familiarity with *a* and *b* words in learning *a_b* dependencies, and the effect instructions might play in learning in two sequential learning phases. I dissociated the effect of prior familiarity from the effect of simply having a pre-familiarization by comparing a Control and an Experimental conditions, both of which presented learners with two-word phrases. In the Experimental Condition these phrases contained the words that would later instantiate remote dependencies at familiarization (*a,b*). In the Control Condition they contained other words that would not be heard again at familiarization or test (*w, v*). I also differentiated between Instruction conditions: while in the Instruction Condition participants were told in advance of the existence of two separate learning phases, in the No Instruction Condition participants were not given this information in advance.

Results showed a near-significant effect of Instruction: across both conditions, participants showed better learning at familiarization when the transition from pre-familiarization was signaled to them in advance. This result is quite surprising if one considers that the only real difference between the two conditions was the awareness participants had of the transition between two learning phases. Participants were not told that the material in the first part would be different from that in the second part. In fact the two learning phases were part of the same 'alien language' as the instructions indicated. Furthermore, participants in the Instructions Condition, being told to click the 'Continue' button after pre-familiarization, proceeded to the next phase immediately, while participants in the No Instructions Condition were given a longer pause to react. Therefore, it is not the pause between the two learning phases but strictly the awareness of the transition between the two phases that facilitated learning at familiarization.

The data showed no other significant effects or interactions. This leaves the research question I address in this study unanswered. I found no clear indication that learning *a_b* dependencies was affected, either negatively or positively, by prior familiarity with the *a* and *b* words. I return to this point in the General Discussion, considering why I may not have found an effect and what would be a better way to investigate such a potential effect.

3.9 General Discussion

In this study I set out to investigate the effect of "Starting Small" in an artificial grammar learning experiment. The premises are two theories that make different claims about the (optimal) starting point of the language acquisition process: "Starting Small" (Newport 1988, 1990; Elman, 1993) or "Starting Big" (Arnon, 2009). Both the "Starting Big" and the "Starting Small" hypothesis generated research looking into what facilitates various types of learning:

identifying the smaller units in the input before identifying the rules that connect them, or starting with larger units or chunks and subsequently breaking these down into their components (Cochran, McDonald & Parault, 1999; Ludden & Gupta, 2000; Kersten & Earles, 2001; Conway, Ellefson & Christiansen, 2003; Arnon, 2009). I set out to extend this research by investigating the role of "Starting Small" in non-adjacent dependency-learning: I asked whether learning the association between remote units in a string benefits from prior familiarity with the individual units, or whether, on the contrary, this prior familiarity blocks the detection of dependencies.

I investigated the effect of "Starting Small" on the learning of an a_iXb_i language in which each of the 3 a_i words co-occurred with only one of the 3 b_i words, while the *X* words varied freely. To indicate learning, participants judged the grammaticality of correct a_iXb_i strings or incorrect a_iXb_j strings in which the string-initial *a* was combined with the wrong *b*-word. An above-chance accuracy in these grammaticality judgments would show that participants developed a sensitivity to the correct $a_i\_b_i$ pairings, and therefore learned the dependencies. In order to assess the role that prior familiarity with the individual *a/b* words had on this learning, in experiment 1 I introduced a learning phase preceding familiarization with the *aXb* language in which learners encountered the *a* and *b* in two-word phrases (*aY* and *Zb*, with *Y* and *Z* sets of bisyllabic nonce words disjunct from the *X* set) from which they could segment and learn them. In 3 different length Conditions, I varied the amount of exposure to the target *a/b* words that learners received prior to familiarization. A comparison between conditions was intended to reveal whether learners allowed to form stronger memory representations for the individual words were better or worse at subsequently learning the associations between them. Contrary to the predictions of the "Starting Small" hypothesis, performance declined with longer pre-familiarization.

Two possible interpretations of these results arose: either prior familiarity with the *a* and *b* words impeded learning of *a_b* dependencies, or else the longer pre-familiarization phases simply induced fatigue, reducing the cognitive resources that the learners could apply to learning the language at familiarization. The second interpretation is a little vague, as we do not know what cognitive resources in particular are engaged during (pre-) familiarization: for instance, recognizing correct non-adjacent dependencies at test has been linked to attention at familiarization (López-Barroso, Cucurell, Rodríguez-Fornells, Diego-Balaguer, 2016), but it is not clear how this attention might decline in time and what amount of pre-familiarization would render learners less able to attend to the input at familiarization. Nevertheless, it was important to dissociate the effect of time/length of pre-familiarization from the effect under investigation, that of prior familiarity with the relevant words.

To achieve this, Experiment 2 investigated the effect of a very short, but more salient pre-familiarization phase. The rationale was that in Experiment 1 a Short Pre-familiarization phase showed comparable learning performance to a condition in which no pre-familiarization was present. I increased the acoustic/prosodic salience of the *a* and *b* elements and assumed this would facilitate their detection and memorization. With these more salient target words, I observed a significant effect of the Short Pre-familiarization: learning was significantly better in the no-pre-familiarization condition than after as little as 1 minute of pre-familiarization.

It is important to note that the interpretation of Experiment 2 is problematic, since it rests crucially on a comparison with conditions tested previously, as part of other experiments. This comparison turned out not to be valid, since a small difference in methodology between Experiment 2 and previous experiments may have crucially affected performance: participants exposed to a Short Pre-familiarization in Experiment 2 were not explicitly instructed about the existence of two separate learning phases, whereas participants in Experiment 1 were. It is also important to note, at this point, that the study presented in this chapter is exploratory in nature: since no previous studies have directly investigated the effect of prior exposure on NAD-learning in an artificial grammar, there is no established methodology for doing so. I investigated the effect in question in different ways across several experiments, identifying the confounds in different experimental designs and working towards an optimal experimental set-up which should capture this effect robustly and without bias. Thus, Experiments 1 and 2 serve mostly as a starting point to understanding the potential confounds and methodological flaws that could obscure the effect we are seeking. They offer no conclusive answer to the research question, but only a baseline for improving the methodology used to address this question.

Experiments 1 and 2 pointed out the importance of dissociating the effect of *having* a pre-familiarization per se from the effect of *learning* from that pre-familiarization. We addressed this in Experiment 3 by directly comparing two conditions that only differed in the *content* of the pre-familiarization. All participants were exposed to a Short Pre-familiarization with the same structure, and subsequently proceeded to the same familiarization and test phase. All participants received the same instructions, and underwent the same procedure. The only and crucial difference was the content of the pre-familiarization: while one group was pre-familiarized with 6 monosyllabic words that would appear in non-adjacent relationships at familiarization, the other group was pre-familiarized with 6 mono-syllabic words that would not occur again at familiarization. A difference in performance between the groups would show the effect of prior familiarity with *a/b* on *aXb*-learning above and beyond the negative impact that any kind of pre-familiarization might have. Furthermore, the pre-familiarization contained not *aX/Yb* phrases but *a/b* words in isolation, ensuring that the pre-familiarization did not direct attention to adjacent instead of non-adjacent regularities, and also that 1-minute of pre-familiarization contained 12 instead of 6 repetitions of each *a/b* word.

The results of Experiment 3 were surprising. After a one-minute exposure to a list of monosyllabic words, learners were unable to detect the rules of the *aXb* language at familiarization – irrespective of whether these words appeared at familiarization or not. These results indicated a negative effect of any type of pre-familiarization on the subsequent learning phase, and suggested that the methodology employed so far is not appropriate for capturing the intended effect.

In Experiment 4, I identified a contributing factor to the negative pre-familiarization effect identified in Experiment 3: the instructions that learners were given prior to the experiment. I compared a condition in which learners were not instructed about the existence of two learning phases, to a condition in which learners were explicitly told about the two phases and instructed to click a button to proceed to the second phase as soon as the first one finished. Although this seemed like a minor manipulation, we obtained a significant effect: the group that was explicitly instructed to press the 'Continue' button after pre-familiarization showed significantly better learning at familiarization.

This result aligns with previous research (Gebhart, Aslin & Newport, 2009; Mitchell & Weiss, 2010; Zinszer & Weiss, 2013) which has shown that participants presented with two different learning phases one after the other only show significant learning in the first phase, unless specific cues mark the transition from one learning phase to the other. It is proposed that once participants tune in to the regularities of the grammar in the first learning phase, their attention to the input starts to decline. If no external cue signals that the input has changed, participants do not recover attention to the second grammar and a primacy effect is observed (only the first grammar is learned).[8]

[8] It is relevant to note that Gebhart, Aslin & Newport (2009) used a 30s pause cue that was sufficient to saliently mark the transition between the two grammars, while in Experiment 4 explicit instructions about the transition were necessary in addition to the pause. In my experiments participants were completing a simultaneous coloring task which may have distracted them, whereas in Gebhart, Aslin & Newport's experiments participants were instructed to pay close attention to the nonsense language they heard. It may be that in incidental learning conditions, more salient cues are necessary to mark transitions between different types of input.

Finally, while controlling for the effect of Instructions, Experiment 4 also tested the effect of prior familiarity with the *a* and *b* words on learning *a_b* dependencies. An Experimental Condition pre-familiarized learners with 36 *aX/Yb* strings, while the Control Condition pre-familiarized another group of learners with 36 similar strings that replaced the target *a/b* words with novel monosyllabic words that would not appear subsequently at familiarization or test. Having eliminated some of the potential confounds identified in previous experiments, I expected a comparison of the Control and Experimental Conditions to reveal the effect I set out to investigate in this chapter. Unfortunately, there was no main effect of Condition and no interaction with Instruction. While the learning effect in the Experimental Condition (with explicit Instructions) was slightly higher than the effect in the corresponding Control Condition, this difference was not statistically significant. The research question I set out to investigate is therefore left unanswered.

Several factors may have contributed to obscuring a potential effect of Condition. Firstly, the exposure phase may not have been sufficient for the learners to acquire stable representations of the target elements *a/b*: participants were exposed for 1 minute to 36 *aX/Yb* strings, where they heard each of the *a/b* elements 6 times, in combination with various other novel words. Because the pre-familiarization phase was meant to immediately precede familiarization, participants could not be tested on their recall of the target words after pre-familiarization. In the absence of this direct measure, indirect evidence of learning at familiarization could be obtained if this learning had an effect on learning at familiarization, namely a comparison of the Experimental and Control conditions which in our experiment yielded no significant difference in performance. Thus, we have no way to assess how far the pre-familiarization phase went in providing the learners with stable memory traces of the target *a/b* words. It is possible that with a larger amount of exposure, the pre-familiarization effect may emerge more sharply in the data.

Secondly, it is unclear what the effect is of having the pre-familiarization immediately precede the familiarization. The phenomenon that I am trying to model in these studies is the gradual transition that children may go through, over several months, from processing smaller units in the input to expanding their STM span to encompass greater chunks and more remote dependencies. Nevertheless, the adults in these experiments are exposed to an abrupt change in the input. As already discussed in the preambles of Experiments 3 and 4, the structure of the input at pre-familiarization may direct learners' attention towards aspects that are salient at pre-familiarization but do not facilitate learning at familiarization. Thus, presenting participants with a pre-familiarization composed of individual words presented in isolation may engage an item-based learning mechanism, whereas presenting them with two-word phrases may prime learners to track local/adjacent relationships. None of these strategies could support NAD-learning at familiarization, but participants may remain committed to them, or may find it effortful to switch to a different learning strategy in real time. Alternatively, participants who were pre-familiarized with the *a* and *b* words and remember these forms may at familiarization shift their attention towards the words they have not heard previously, namely the intervening *X* words, and therefore be less likely to note the new *a_b* dependencies (see Soderstrom, White, Conwell & Morgan, 2007, for evidence that infants have difficulties tracking dependencies between familiar function words over unfamiliar (nonce) but not familiar words). Finally, it is possible that participants who have learned at pre-familiarization are simply fatigued, and have fewer cognitive resources to attend to a different type of input at familiarization.

It is worth noting that a Control Condition with a pre-familiarization that is entirely unrelated to the subsequent familiarization and test should show a learning performance similar to a condition in which no pre-familiarization is present, if the existence of a pre-familiarization itself has no negative effect on learning. In Experiment 4, the Control Condition with explicit Instructions showed a somewhat smaller learning effect than in the No Pre-familiarization Condition in Experiment 2. From these findings we cannot conclude for certain whether this difference is accidental or speaks to an enduring negative effect of pre-familiarization, and we can only speculate as to the possible ways in which the immediately preceding pre-familiarization may affect learning at familiarization. Nevertheless, we cannot rule out the possibility that an immediately preceding pre-familiarization phase may affect learning at familiarization, and in order to avoid threats to internal validity future research must eliminate this confound.

To address these potential issues which may have obscured the true effect of prior familiarity in our experiments, I propose that future research should adopt a more suitable design, separating the two learning phases (pre-familiarization and familiarization) in two consecutive days. Participants on the first day should be exposed to a longer pre-familiarization with the words of the subsequent *aXb* language, and subsequently do a short recall test to have a direct assessment of the degree to which they memorized the relevant words. In this way, participants would commit the words to long-term memory. On the second day, participants would be exposed to the familiarization phase separately, with no prior distractions or fatigue, and with no difficulty keeping the two learning phases separate. They will identify (some) of the words of the *aXb* language with words they recall from the previous exposure phase, in much the same way an infant may identify familiar words in the language she hears, thus ensuring the ecological validity of this experimental design. Following this, they would undergo the test phase where they will judge the grammaticality of novel *aX'b* strings.

As a final remark, it is important to point out how the results in this chapter may reflect on prior research. Arnon & Ramscar (2012) found a significant inhibitory effect of prior familiarity with individual words on learning (adjacent) regularities between them. However, their study did not introduce an explicit separation between the two subsequent word-learning and rule-learning phases. The current findings on the relevance of the primacy effect in an incremental learning task suggest that Arnon & Ramscar's results could have been entirely determined by this confound. Participants in their experiment might have learned during the first phase of training but lost their attention to the stimuli during the second stage of training. This could explain why participants exposed to individual words did not learn regularities between them subsequently, but participants first exposed to the regularities learned both the words *and* the regularities, since both were present in the initial familiarization phase. In Experiment 4 participants exposed to target words in short phrases were not inhibited (but in fact showed a non-significant advantage) in learning NADs between these words subsequently, as long as they were explicitly informed of the transition between pre-familiarization and familiarization. In Arnon & Ramscar's (2012) model, these learners should have focused on associating the *a/b* elements with the predictive cues that were strongest at both pre-familiarization and familiarization (the positional cues, i.e. the fact that *a* was string-initial and *b* string-final), and these associations should have blocked the association between specific *a* and *b* tokens which was only present at familiarization. The fact that this did not happen is reason to be skeptical of Arnon & Ramscar's (2012) results and conclusions.

To summarize this chapter, I have attempted to answer the research question of whether prior familiarity with a set of words supports or hinders the subsequent detection of remote dependencies between those words. A series of experiments have been run that yielded conflicted findings, suggesting that the transition from simple to complex input is difficult to model in laboratory conditions. I believe the data presented in this chapter should serve to highlight the confounds that can arise with this type of experimental set-up and to inform a better methodology to investigate the research question. I therefore consider this an exploratory study, with no conclusive results, but with a series of suggestions for tackling a difficult and elusive question on the nature of NAD-learning and the intricate dynamics between item-based learning and rule learning.

Chapter 4

Generalization and Prosodic Cues to Non-Adjacent Dependency-Learning in Infants

In previous chapters I have shown that adult learners are capable of detecting remote dependencies in spoken input, and of recognizing them in novel contexts. I have shown that NAD-learning is sensitive to prosodic cues, and that it may be guided by Gestalt principles of perception: dependencies are more easily detected in contexts where the dependent elements are prosodically distinct (and similar to each other).

In what follows I turn my attention to infants. Distributional learning mechanisms have been studied both in adult and infant populations, with the result that the learning abilities found in adult populations were also found in infant populations. Thus, both adults (Saffran, Newport & Aslin, 1996) and infants (Saffran, Aslin & Newport, 1996) are able to segment words from the input based on the transitional probabilities between syllables, as well as use distributional cues to infer phonetic categories (Maye & Gerken, 2000; Maye, Werker & Gerken, 2002). Both infants and adults have also been shown to learn finite-state grammars and apply those grammar rules to novel stimuli (Gómez & Schvaneveldt, 1994; Gómez & Gerken, 1999), use frequent frames in the input (*to___it*) to infer lexical categories (*to V it*; Mintz 2002, 2006), detect the adjacent dependencies that determine phrase structure (Saffran, 2001), and the non-adjacent dependencies that indicate more remote syntactic relationships (Gómez, 2002; Gómez & Maye, 2005). In sum, a large body of work has shown similar learning abilities in infants and adults, while fewer studies have reported differences between these two groups. One specific area where adults and infants differ is in the way they deal with noise and irregularity in the input. Unlike adults, infants do not faithfully reproduce the variability in the input, but are prone to over-regularization (Hudson Kam & Newport, 2005, 2009; Austin & Newport, 2011). Austin & Newport (2011), for instance, exposed both adults and infants to a miniature language where the same pseudo-nouns were preceded 67% of the time by one pseudo-determiner and 33% by another. While adults reflected this probabilistic distribution faithfully by using the less frequent determiner 33% of the time in their own productions, infants overgeneralized their use of the frequent determiner and almost never produced the less frequent one.

Apart from this line of research investigating infant and adult tendencies to overgeneralize, however, few studies have actually achieved a direct comparison of learning in adult and infants. Behavioral Artificial Language Learning experiments employ fundamentally different methodologies for infants and adults. While adult learning can be assessed through direct tasks like choosing correct ('grammatical') over incorrect ('ungrammatical') strings, or reproducing strings of the language (or even measuring reaction times to the *b* element in an *aXb* string when it is correctly predicted by *a* vs. when it is not, see Misyak, Christiansen & Tomblin, 2010), infant learning is assessed indirectly, in particular by measuring the amount of attention they show to (novel) language-consistent versus language-inconsistent strings. Furthermore, infants have limited attentional resourcescompared to adults. This means infant experiments are typically brief (only a few minutes of exposure from which the infants need to acquire the regularities of the AG), and the drop-out rate due to fatigue, fussiness and lack of attention is still quite high. Therefore, infant behavioral experiments are short and employ highly simplified materials, while adults can be presented with more complex input for a longer period of time. Note for instance that Gómez (2002) exposed adult participants to 432 strings containing 3 *a_b* dependencies for 18 minutes, while she exposed 18-month-olds to 48 strings containing 2 *a_b* dependencies during a 3-minute training phase.

If behavioral studies fall short of directly comparing infant and adult learning, neurophysiological studies may be a more promising avenue, as they can be used in the same way with adults *and* infants. Furthermore, they do not merely show evidence that there *is* learning in an AGL condition, but also identify the physiological correlates of learning which can (arguably) be compared between different populations. For instance, two EEG studies, Mueller, Oberecker & Friederici (2009) and Friederici, Mueller & Oberecker (2011), investigated learning of non-adjacent dependencies in Italian (*La sorella* ***sta*** *cant**ando***, ‘The sister **is** sing**ing**’) in adults and infants, respectively. German 4-month-olds familiarized with the Italian input exhibited a similar late positive (centro-parietal) ERP response when presented with strings that violated the familiarized dependency (**La sorella* ***sta*** *cant**are***) just like Italian native adults, whereas German adult learners exhibited a very different (frontally-distributed) positivity. In this, all groups showed discrimination between grammatical and ungrammatical strings, but because German infants and native Italian speakers patterned together with respect to their responses, whereas German adults showed a different response, the authors concluded that infants (unlike adults) could develop a more mature, native-like representations of a grammar even after a brief amount of exposure. However, this claim needs to be explored more in depth in further studies, replicating and extending these results.

In this chapter I investigate similarities between adults and infants in an indirect way. That is, I inquire if the NAD-learning mechanism is constrained in similar ways in both groups. In Chapter 2 I showed that adult learners can detect co-occurrence patterns between non-adjacent words in a string (*aXb*), and demonstrate a preference for those patterns even when encountering them in novel contexts (*aX'b*). In this chapter I ask whether infants show a similar power of generalization. Previous studies have shown that infants are capable of generalizing the patterns they learn to novel contexts from a very young age (Marcus, Vijayan, Bandi Rao & Vishton, 1999; Gómez & Gerken, 1999; Gerken, Wilson & Lewis, 2005) but none so far have tested this generalization ability with respect to NAD-learning. However, NAD-learning studies have suggested that abstracting away from the context in order to attend to the non-adjacent relationships is a crucial factor in detecting dependencies (Gómez, 2002; Gómez & Maye, 2005). Experiment 1 tests the prediction that 18-month-old infants can generalize NADs by detecting previously learned dependencies in novel strings, with unfamiliar intervening elements.

Furthermore, Chapter 2 also showed that adults can detect dependencies between either prosodically marked or prosodically reduced elements, as long as these are distinct from the intervening material. In that chapter I proposed that they are guided by Gestalt principles of perception, which allow them to detect patterns between items that are perceptually similar to each other but distinct from the context. Experiment 2 investigates whether 18-month-olds are also capable of detecting dependencies between prosodically reduced elements in an ALL experiment.

4.1 Experiment 1

4.1.1 Introduction

In Chapter 2, adult learners showed the ability not just to identify the patterns between non-adjacent words in the input, but also to generalize those patterns to novel contexts. This ability seems crucial for a learning mechanism that is meant to aid learners to acquire morpho-syntactic dependencies as in *The princess* ***is*** *lovingly kiss****ing*** *the frog*, since these morpho-syntactic rules must apply irrespective of the sentential context in which they occur (*The king* ***is*** *angrily glar****ing*** *at the prime-minister*). However no Artificial Grammar Learning studies so far have attempted to test this ability to generalize NADs to novel contexts. In this chapter, therefore, I attempt to replicate our findings from Chapter 2 with infants, and show that they too possess the ability not just to detect non-adjacent patterns in the input, but also to generalize these patterns to novel strings.

Infants as young as 7 months have been shown to derive abstract representations of same/different patterns in short strings and generalize them to novel stimuli. Marcus et al. (1999) showed that 7-month-olds familiarized with trisyllabic strings with the structure ABB (e.g. *ga ti ti*, *li na na*) showed a subsequent preference for novel ABB (e.g. *wo fe fe*) versus ABA (e.g. *wo fe wo*) strings at test, even though they had never heard either of them before. Infants familiarized with ABA (e.g. *ga ti ga*, *li na li*) strings showed the exact opposite preference.[9] Similarly, Gómez & Gerken (1999) exposed 11-12 month-olds to a finite-state grammar, where sensitivity to adjacent probabilities was crucial to detecting the structure of the language. In one of their experiments infants were familiarized with a set of strings generated by that grammar (e.g. *fim jed tup fim jed tup*) and subsequently showed a preference for completely novel strings that were consistent with the grammar (e.g. *pel rud jic pel rud jic*) over novel strings that weren't (e.g. *pel tam pel pel pel jic*). These findings indicate an early ability to develop abstract representations of structure that are not linked to specific stimuli, but can be applied to different sets of stimuli.

A subsequent study (Gerken, 2006) showed that infants could develop either abstract *or* stimulus-specific representations, depending on the input. When children were familiarized with languages similar to Marcus et al.'s (1999), with both A and B classes showing high variability, they generalized ABA or AAB to novel stimuli. However when the B class was represented by only one stimulus, the syllable *di*, infants did not abstract the general pattern ABA or AAB from their input, but rather the item-bound representations A*di*A or AA*di*. In other words, when exposed to strings like *ga di ga* and *li di li*, 9-month-olds showed a preference for *po di po* over *po po di*, but not a preference for *ba ko ba* over *ba ba ko*.

Aslin & Newport (2012) proposed that the findings of Marcus et al. (1999) and Gerken (2006) can be construed as evidence that abstract generalization and item-based/stimulus-specific learning are in fact two manifestations of the same basic statistical learning mechanism. According to Aslin & Newport, the same statistical learning mechanism can compute co-occurrence patterns between either abstract representations or specific item representations. The nature of the representations that undergo statistical computations depends on the variability in the input. When the input presents high variability, learners will find it difficult to fully encode individual items, and will instead encode them only superficially along a few salient dimensions (e.g. CV syllable, high/low vowel, etc.).

[9] Note, however, that subsequent studies showed difficulties replicating and extending the findings of Marcus et al. (1999) (Geambasu & Levelt, 2015), suggesting that generalization in infancy may not be very robust.

Furthermore, when a variety of items occur interchangeably in the same context(s), these will be grouped together in a class, such that statistical computations can apply over classes of items. However, when items can be fully encoded because they are frequent in the input or have become sufficiently familiar to the learner, sensitivity to item-specific patterns can arise.

Aslin & Newport's (2012) proposal goes against the claim that item-based and rule-based learning are separate mechanisms (Endress & Bonatti, 2006). The advantage of their one-system account for statistical learning is not just parsimony, but also the fact that no additional explanations have to be offered for when and why learners switch from abstract generalizations to stimulus-specific learning. It is not the learner that 'decides' between drawing generalizations or computing statistics between items, but it is the (variability in the) input that determines the type of representations over which the learner computes co-occurrence statistics.

The notion that variability can induce generalization has also been proposed in connection to NAD-learning. Gómez (2002) and Gómez & Maye (2005) showed that both adults and infants acquired the *a_b* dependencies in *aXb* strings better with higher variability in the class of intervening *X* items. Gómez (2002) claimed that the mid-string variability might have helped focus attention on the more stable elements, the dependent *a* and *b* words, instead of the variable *X*s (and implicitly on the higher co-occurrence probability between the non-adjacent words instead of the lower co-occurrence probabilities between the adjacent words). Under Aslin & Newport's (2012) hypothesis, increasing the variability of the *X* elements amounts to obtaining an abstract representation of the form *a_b* in which the representation of the middle element is underspecified and contains only a handful of salient features that all *X* items have (bisyllabic, trochaic stress pattern, etc.). Thus, if 18-month-olds learn remote dependencies more easily when the intervening *X* element is more variable, then these infants must be forming semi-abstract representations in which the dependencies can combine productively with any number of intervening *X*s.

Although the prediction by Aslin & Newport (2012) and Gómez (2002) seems straightforward, to my knowledge no published studies so far have specifically investigated infants' ability to generalize NADs to novel contexts. Previous studies have shown that by 18 months (and as early as 12 months) infants are capable of abstracting multiple categories from the input (*X* and *Y*), identifying combinatorial patterns between those categories and grammatical markers (*a* combines with *X*, *b* combines with *Y*) and generalizing the combinatorial properties of a category to novel potential members of that category (novel *X'* combines with *a,* not *b*; Gómez & Lakusta, 2004; Gerken, Wilson & Lewis, 2005). Furthermore, 12-month-olds infants acquiring *aX* adjacent dependencies (between a grammatical marker *a* and a category *X*) can subsequently generalize these to non-adjacent *acX* dependencies in which a novel *c* is inserted into the dependency (Lany & Gómez, 2008). This suggests that infants do not simply detect dependencies but are also capable of generalizing those dependencies to novel contexts. In short, I expect that 18-month-olds should be able to use *aXb* strings to abstract the category containing variable *X*s and extract generalizable *a_b* frames.

In this experiment I expose infants of 18 months to an artificial *aXb* language and test their acquisition of the *a_b* dependency with novel *aX'b* strings. Based on the findings above, I predict that infants with a mature ability to detect NADs in an unfamiliar language (i.e. most infants at the age of 18 months) should also be able to generalize those dependencies to novel contexts. I employ the same methodology as used in previous NAD-learning studies with infants (Gómez, 2002; Gómez & Maye, 2005; Kerkhoff et al., 2013), but only slightly adjust the test phase to address the unexplored research question of whether 18-month-old infants can generalize the NADs they acquire to novel environments (i.e. *aX'b* strings with novel *X'*s). The current experiment is also intended to be comparable to adult learning in the Emphasized 250 ms pause condition (Chapter 2).

4.1.2 Materials

Subjects were exposed to a language *aXb* similar to Gómez (2002), with stimuli adapted to Dutch taken from Kerkhoff et al. (2013). The language contained 2 $a_i\_b_i$ dependencies (*tep/sot__lut/jik*) and 24 bisyllabic intervening *X*s (*wadim, kasi, poemer, kengel, domo, loga, gopem, naspu, hiftam, dieta, vami, snigger, rogges, densim, fidang, rajee, seta, noeba, plizet, banip, movig, sulep, nilbo* and *wiffel*), combined exhaustively into a total of 48 familiarization strings. The *X* set size of 24 is the largest set size used previously in this type of experiment (Gómez, 2002; Gómez & Maye, 2005; Kerkhoff et al., 2013) and should therefore be optimal not just for prompting NAD-learning but also for inducing generalization over the intervening *X* class.

Two Language Versions were created (L1, L2), which differed only in the $a_i\_b_i$ pairings (L1 had the dependencies *tep_lut* and *sot_jik*, whereas L2 had the dependencies *tep_jik* and *sot_lut*), such that the strings that were grammatical to L1 were ungrammatical to L2 and vice-versa. Infants were randomly assigned to one or the other Language Version, to control for stimulus biases. Because I wanted to investigate children's ability to generalize the dependencies learned to novel contexts, I created test strings employing three *novel X*s (*klepin, lotup,* *tarsin*). Each of the three *X*s was combined with the dependencies of L1 and of L2, to create novel test strings that were either consistent with L1 or with L2.

The stimuli were recorded in a sound attenuated booth, at 48kHz, using a TASCAM DA-40 DAT-recorder. A female speaker read out *aXb* strings in a child-friendly voice, laying special emphasis on the *a/b* elements. The string-initial *a* and string-final *b* tokens were spliced from these recordings, in replication of the Emphasized Condition in the adult study in Chapter 2. Also, in order to replicate the properties of the stimuli in that Condition, I recorded the *X*s separately, in carrier sentences in Dutch where they occupied the slot of the direct object noun:

(1) *Ik zie de ____ in de tuin.*

I see the __ in the garden.

e.g. *Ik zie de* ***wadim*** *in de tuin*.

Previous studies (Gómez, 2002; Gómez & Maye, 2005; Kerkhoff et al., 2013) recorded *aXb* strings read out in a child-friendly (motherese-type) manner, and spliced both the *a/b* and the *X* elements from these recordings. However, as this series of experiments is meant to replicate our findings with adults, it was desirable/necessary to create the stimuli in the same way as in the adult experiments in Chapter 2 (Emphasized Condition, 250ms). Acoustic properties of the *a, b* and *X* tokens, as analyzed using Praat 5303, are presented in Table A4.1 in the Appendix, alongside the acoustic properties of tokens employed in Kerkhoff et al. (2013). As can be seen, in both sets of stimuli the *a/b* tokens have a higher mean pitch than the *X* tokens, rendering them more salient.

All the tokens were spliced from the original recordings and concatenated into *aXb* strings with 250 ms within-string pauses between nonce words. Strings were approximately 2s in duration and were played separated by 750ms pauses.

4.1.3 Participants

Infants were recruited through written invitation to parents, whose addresses were provided by the Utrecht municipality. A total of 31 infants were included (16 females, 15 males), with an average age of 18 months and 16 days (range: 18 months and 4 days - 19 months). All infants included had normal birthweight (2500-4500 grams), were not pre- or post-term (had a gestation period of 35-42 weeks), and had no known neurological, hearing or vision problems.

An additional 34 infants were tested but not included, due to: low birthweight (*n* = 3), oxygen shortage at birth (n = 1), vision impairment (n = 1), failure to retrieve information about birthweight and gestation period from parents (n = 2), fussiness, restlessness, crying or fatigue (n = 18), completing fewer than 2 valid consistent or 2 valid inconsistent trials (n = 6), or technical error (n = 3). The drop-out rate due to infant behavior (fussiness, crying, etc.) is 28% (18/65), which is comparable to previous studies (32% in Kerkhoff et al., 2013 and 29% in Gómez, 2002). Exclusion due to short looking times (completing fewer than 2 valid trials of each type) is also similar to previous studies (6/65, 9% the same as in Kerkhoff et al. 2013).

From the total of 248 test trials of the 31 infants included (8 test trials per infant), 35 were excluded because they totaled a looking time shorter than 2 seconds. An additional 4 trials were excluded due to experimenter error (the trial was ended too soon), leaving 209 valid trials (84.2%) to be analyzed.

In addition to providing details about the infants' birth weight, gestation period, and medical history, parents also completed the Dutch version of the MacArthur-Bates Communicative Development Inventory (N-CDI; Zink & Lejaegere, 2002), to establish receptive and productive vocabulary size at 18 months.

4.1.4 Procedure

I employed the Headturn Preference Procedure (cf. Kemler-Nelson et al., 1995), similar to previous NAD-learning studies (Gómez, 2002; Gómez & Maye, 2005; Kerkhoff et al., 2013). The Headturn Preference Procedure relies on the assumption that when infants discriminate structurally consistent from inconsistent stimuli they will preferentially attend to one type of stimulus over the other. Furthermore, it is also assumed that attention to auditory stimuli can be approximated by visual attention – duration of visual fixation – of a visual stimulus associated with a particular auditory stimulus (Kemler-Nelson et al., 1995).

Infants were tested individually, while seated on their caretaker's lap in a sound-attenuated booth. Throughout the experiment caretakers listened to music over a set of headphones, ensuring that they were unaware of the stimuli being played and thus could not bias infants' behavior in any way. The stimuli played over two speakers located on either side of the infant, at eye-level. In front of each speaker was a red light, consisting of three concentric rows of LEDs that lit up sequentially, while a third, green light of the same description was positioned on the wall directly in front of the infant. A camera was mounted directly above the green light and recorded the looking behavior of the infant. An experimenter in the adjoining room viewed the camera feed on a monitor and controlled the lights in the cabin. All parents provided written consent that the video recordings could be used to analyze the data, and that the data could be used for publication.

Familiarization The green light in front was used to capture the infant's attention, and as soon as the infant's gaze oriented towards it, the green light was extinguished and one of the red lights on the side began blinking, while simultaneously the sound began playing over both speakers. When the infant looked away from the side light for more than two seconds, this light was extinguished, and the front green light was lit again. The infants' looking time to the blinking side light was measure from the moment she oriented towards the light and until the moment she looked away from more than two seconds. As soon as the infant oriented towards the green light it was extinguished, one of the side-lights was turned on again and the procedure was repeated. The successive use of one or the other red side-light was quasi-randomized so that the same side-light was not employed more than twice in succession.

The familiarization language consisting of the 48 *aXb* strings (presented in random order) was played continuously over both speakers, irrespective of the infant's looking behavior. This was done in order to ensure that the familiarization phase had the same duration for all infants (two minutes) and was not prolonged by the infants' failure to visually attend to the stimuli. However, during this phase infants' looking times (the total amount of looking time to the red side-lights during familiarization, or TLTfam) were still calculated, to obtain an approximate measure of each infant's attention to the familiarization material, on the assumption stated above that visual fixation is a proxy for auditory attention.

Contingency Phase Following the familiarization phase, infants were presented with a short contingency training phase. The green light was used to get the infant's attention, and as soon as the infant oriented to it, one of the red lights started blinking. When the infant oriented to the red side-light, a pure tone of 440Hz lasting 1s was played repeatedly (with 125ms pauses in between) from the speaker on the same side as the light for as long as the child fixated the red light. When the child looked away for more than two seconds the sound stopped, and the green light came on again (whereas if the child looked away for less than two seconds, the trial continued uninterrupted). The same procedure was repeated one more time with the opposite side-light. This two-trial contingency phase was intended to train the infants on the contingency of light and sound on their looking behavior, thus preparing them for the procedure that would be used subsequently in the test phase and reducing potential noise in the first few test trials due to infants familiarizing themselves with this procedure. Contingency training has been used successfully in previous infant experiments (Mintz 2006, 2013) but has not been used in NAD-learning experiments before (Gómez, 2002; Gómez & Maye, 2005; Kerkhoff et al., 2013).

Test Phase The test phase consisted of 8 trials, 4 consistent with L1 and 4 consistent with L2. The order of presentation was quasi-randomized, such that no more than two trials of the same type (L1- or L2-consistent) could occur simultaneously. Whether the first trial was consistent or not to the familiarization language of the infant was counterbalanced over infants. Each test trial contained a maximum of 15 *aX'b* strings (played consecutively with 750ms pauses in between), randomly sampled from the 6 *aX'b* combinations (3 novel *X'*s combined with 2 dependencies) created for each language, L1 or L2.

The procedure was the same as in the contingency phase. For each trial, as soon as the green light captured the infant's attention, the red light started blinking. When the child oriented to the red light the stimuli started playing, until the child looked away for more than two seconds and the trial stopped. Children's looking times to the red light was measured in each trial. In order to obtain more precise measurements, I recoded all the video recordings offline, using PsyCode, a Mac application that allows the precise identification of look or look-away points in the recording on a frame-by-frame basis (provided to the UiL-OTS labs courtesy of Judith Gervain and Luca Bonatti).

4.1.5 Results

Table C4.1 shows the mean looking times (across trials and across infants) to test trials consistent or inconsistent with the language of exposure as well as the mean N-CDI scores (raw numbers of words produced / understood, as well as percentile scores), and the total looking time at familiarization, quantifying the amount of attention infants had during exposure to the language.

As can be seen, infants in our sample varied greatly, both in terms of language development (receptive and productive vocabulary) and in terms of their attention at familiarization. The mean N-CDI percentile scores for receptive and productive vocabulary were around 60, comparable to those reported by Kerkhoff et al. (2013).

	Mean (Standard Deviation)	Range
LT Consistent	M = 10.04 s (SD = 4.87) Median = 5.76 s	3.72 – 24.27 s
LT Inconsistent	M = 11.42 s (SD = 7.23) Median = 7.7 s	3.92 – 35.65 s
Productive vocabulary	Raw M = 76.7 words (SD = 63.62) Percentile M = 59.79 (SD = 27.04)	0-259 words 1st – 95th perc.
Receptive vocabulary	Raw M = 257.2 words (SD = 102.95) Perc. M = 65. 96 (SD = 27.47)	29 – 440 words 1st – 95th perc.
LT at Familiarization	M = 62.328 s (SD = 15.03)	31.11 – 88.29 s

Table C4.1. Results for Experiment 1

Previous studies (Gómez, 2002; Gómez & Maye, 2005; Kerkhoff et al., 2013) used Repeated Measures ANOVAs to analyze this type of data, averaging looking times for consistent and inconsistent stimuli across trials. However, because they average looking times across trials, RM ANOVAs do not take into consideration the variation within the different (consistent or inconsistent) trials that may arise within a subject. That is, a child may look longer to some trials of the same type than others, depending on a variety of factors such as whether these trials occur later or earlier in the test phase. RM ANOVA takes this variation to be equal to 0, and assumes there are only 2 observations available per infant (mean looking times to consistent and to inconsistent stimuli) where in fact there are (up to) 8. RM ANOVA is therefore not a sufficiently conservative test – it assumes the data is less noisy than it actually is.

Linear Mixed Models can take into consideration variation both at the level of the Subject (differences between participants) and Trial Number (looking times differences across trials). In addition, Linear Mixed Models are robust with respect to missing data, which is particularly suited to infant datasets where short trials are excluded. I therefore used a Linear Mixed Model with Looking Times at test (LT, per participant per test trial) as the (continuous) dependent variable and Subject as random intercept. I also introduced Trial Number (1-8) as factor to control for differences between, for instance, earlier trials where the infants may show greater attention to the stimuli and later trials where infants may lose visual attention. Because LTs (across trials and participants) followed a logarithmic-like distribution (with a high frequency of short LTs and a low frequency of longer LTs), I used a log-transformation of the dependent variable LT.

I used Trial Type (Consistent or Inconsistent) and Language (L1, L2) as fixed factors, and Trial Number as covariate. I started from a model with Subjects as random intercept and Trial Type (the variable of interest) as fixed factor, and introduced each fixed factor/covariate separately in the model. I used a likelihood-ratio test to compare between models. Language did not significantly improve the model, but Trial Number did. There was a near-significant effect of Trial Number (F (1, 209) = 3.072, p = .081): infants' attention seemed to decline across trials (suggesting loss of attention due to fatigue), with shortest looking times for the last (8^{th}) test trial. There was no significant effect of Trial Type (p = .375).

I also wanted to see whether the nature of the first test trial influenced behavior during the subsequent test trials. Gómez, Bootzin and Nadel (2006) showed that infants who napped between the familiarization and test phase of a NAD-learning experiment used the first trial of the test phase to guide their preference for consistent or inconsistent stimuli during the subsequent test trials (see also Hupbach, Gómez, Bootzin & Nadel, 2009). In the current experiment, infants did not nap, but a contingency training phase did separate familiarization from test. Furthermore, because the stimuli at test were novel, the first test trial may be crucial in re-familiarizing the infants with the dependencies they were meant to track, in their novel contexts. Therefore, I ran the Linear Mixed Model again, excluding LTs for the first trial from the data, and including First Trial (Consistent vs. Inconsistent) as a fixed factor (similar to Gómez, Bootzin and Nadel, 2006). The best model included Trial Type and First Trial as fixed factors and Trial Number as covariate. This model showed a near-significant effect of First Trial (F (1, 280) = 3.639, p = .058) and no other effects. Thus, infants who heard a first trial consistent with the familiarization subsequently showed longer looking times overall than infants who heard an inconsistent first trial.

Importantly, I did not find a significant interaction between First Trial and Trial Type. Infants did not show a significant preference in trials 2-8 for the type of stimulus they heard during the first trial[10].

[10] For all models run for this first Experiment, the final Hessian Matrix was not positive definite, prompting caution in interpreting the results. One possible reason is the introduction of subject as random intercept: the Hessian Matrix can only be positive definite if there is variation in intercepts between subjects (looking times vary between more than within subjects). It is possible that in the current data the random intercept was similar for all subject. I therefore ran all the analyses reported here excluding subject as

4.1.6 Discussion

In Experiment 1 I set out to investigate infants' ability to generalize the patterns they have observed in the input to novel contexts. Infants were exposed to 48 *aXb* strings, with two *a_b* dependencies and 24 different *X* items. Therefore, while they heard each of the dependencies they were meant to acquire 24 times, they only heard each of the 24 intervening *X* elements twice. According to previous research (Gómez, 2002; Gómez & Maye, 2005), this high variability and low frequency of the intervening material should be beneficial to the detection of *a_b* dependencies. Furthermore, previous findings show that 18-month-olds are capable of forming categories between sets of elements occurring in the same context, and of generalizing to novel elements (Gómez & Lakusta, 2004; Gerken, Wilson & Lewis, 2005). I therefore predicted that the high variability and low frequency of the intervening *X* element should prompt young learners not to develop item-specific representations of the *aXb* strings, but rather to develop a quasi-abstract *a_b* frame where any *X* element (of a quasi-abstract category *X*) can be inserted. To investigate that, I tested infants' discrimination of novel $a_iX'b_i$ from $a_iX'b_j$ strings at test.

The results, surprisingly, yielded no learning effect in the group of infants tested. In what follows I turn my attention to two potential reasons why this effect was not observed: (i) the possibility that the methodology employed in this experiment was not appropriate for capturing the predicted effect, or (ii) the possibility that generalization of NADs may still be limited at the age of 18 months.

The current study follows very closely previous research from two different labs, Rebecca Gómez's (Gómez, 2002; Gómez & Maye, 2005) and our own lab at UiL-OTS (Kerkhoff et al., 2013). The finding that 18-month-olds are capable of detecting remote dependencies in simple *aXb* strings has therefore been replicated a number of times, including with Dutch infants in the same facility where the current study was carried out. The methodology employed here is identical to methodology used in these previous studies (the same Headturn Preference Procedure, same selection criteria for the participants, comparable drop-out rates, etc.) in all but two specific aspects: the way the stimuli were recorded, and the introduction of a contingency training phase between the familiarization and the test phase.

The stimuli, as discussed in section 4.1.2 were recorded in the same way as the stimuli for the adult experiments in Chapter 2. While the *a/b* tokens were

a random intercept: the final Hessian Matrix was positive definite and the results did not differ from the ones reported here.

recorded in the same way as in the previous studies of Gómez and Kerkhoff et al., the *X* tokens were recorded in carrier sentences in the slot of direct object nouns. Table A4.1 in the Appendix shows the acoustic properties of the stimuli I used compared to the stimuli used in Kerkhoff et al. (2013). As can be seen, the stimuli were comparable in duration and intensity, but not in pitch – the stimuli in the current study were higher-pitched, particularly the *a/b* tokens. An important consequence of the way stimuli were recorded was that in the Kerkhoff et al. study *X* items had lower mean pitch over their first syllable (despite being stress-initial words). Because the *aXb* strings were read out in a child-friendly voice, the reader emphasized the *a/b* words and de-accented the word in between. The *X* elements in the current study were recorded in contexts where they were in a position of focus/prosodic prominence (direct object nouns preceded by a prosodically reduced functional word). Their first syllable had higher mean pitch that the overall words.

There is no *a priori* reason to assume that the de-accented first syllable of the *X* token facilitated the extraction of the *a_b* dependency in the Kerkhoff et al. study. In that study and the present one, *a/b* tokens still had higher mean pitch than the *X* tokens (first syllable or overall), and were thus highly salient. One possibility, however, is that the stimuli in Kerkhoff et al. were more *natural* in their prosodic contour since the different words were all recorded in exactly the context they would appear in (i.e. at the beginning, middle or end of an *aXb* string), whereas in this study *X* tokens were excised from a completely different context, a procedure which obviously had an effect on their phonetic properties.

There are reasons to be skeptical about this difference as an explanation for the current results. In Chapter 2 adult participants were exposed to stimuli recorded in exactly the same way as here (the Emphasized Condition of Experiment 1), and showed successful (incidental) learning of the dependencies. Furthermore, adult learners failed to learn the *aXb* language when the stimuli did not have a prosodically natural contour (the Functional Condition in Experiment 1), but showed an improvement in learning as soon as the prosodic contour was amended (Experiment 2). Thus, adults are very particular about the type of stimuli they can learn from – in particular with respect to prosodic contour – but they showed robust learning with stimuli similar to those used in the current experiment. This raises the question of why infants should not be able to learn from the same type of stimuli.

It is nonetheless difficult at this point to eliminate the possibility that the stimuli employed in this experiment hindered learning of the remote dependencies. In Experiment 2 of this chapter I employ a different set of stimuli, which follow the prosodic contour of a functor-lexeme-functor sequence in natural languages (such as the determiner + noun + diminutive suffix in Dutch: ***het*** *zebra'**tje***, 'the little zebra'). Here all stimuli will be recorded in the same type of context, and will exhibit the natural prosodic contour of a noun phrase in Dutch, thus eliminating the potential confound of Experiment 1. If 18-month-old Dutch infants can apply NAD-learning successfully to morpho-syntactic dependencies in natural languages (like the *het_je* dependency in Dutch), then they should show significant learning in Experiment 2.

A second methodological difference between Experiment 1 and previous NAD-learning studies is the introduction of a contingency training phase before the test phase. It is unclear why this contingency phase should have any negative effect on learning. It is brief, containing only two trials, and is in fact intended to reduce the noise in the data by familiarizing infants with the procedure of the test trials. It has been used successfully in experiments using the Headturn Preference Procedure (Mintz 2006, 2013). One potential disadvantage is that it prolongs the experiment, increasing the possibility of loss of attention towards the end of the experiment –which might be linked to the change in looking behavior during the last 3 trials of the test phase. However, the contingency phase never lasted longer than 30 seconds. Furthermore, if the contingency phase increased fatigue, we would expect it to also prompt higher drop-out rates due to fussiness or insufficient valid trials. Instead our drop-out rates were comparable to previous NAD-learning studies that did not use this contingency phase.

Finally, while we cannot eliminate the possibility that small methodological changes may be responsible for the failure to obtain the expected learning effect in Experiment 1, we must also consider the possibility that generalization may still be difficult for 18-month-olds acquiring NADs under the conditions of this experiment. As pointed out previously, by 18 months infants are capable of generalizing over classes of items, and this generalization may arise as a result of high variability and low item frequency. Indeed, in this experiment infants heard as many as 24 different *X* items, each of which was iterated only twice. It is difficult to believe that infants formed item-specific representations for each (or most) of the *X*s, so that they memorized specific *aXb* strings instead of abstracting the general *a_b* frame. Indeed, this strategy would run counter to Gómez (2002) and Gómez & Maye's (2005) findings that children benefit from the high variability and low frequency of the intervening *X*.

An alternative explanation is that infants did not fail to learn the dependencies but rather failed to attend to them at test. Previous studies have shown that even when infants are familiar with remote dependencies, they may have a hard time recognizing them in certain contexts, especially where working memory demands are high (Santelmann & Jusczyk, 1998). One study in particular (Höhle et al., 2006) claimed that dependency-recognition can be blocked when the intervening material is not easily identifiable and quickly processed. In the present study, because the intervening *X'*s in the test phase were novel, infants may have taken longer to categorize (or process) them; because of these higher processing demands of the novel *X'* items, upon reaching the string-final *b* element learners might have already lost track of the string-initial *a* element, therefore failing to detect whether the remote dependency was in fact grammatical.

Furthermore, while Gómez (2002) and Gómez & Maye (2005) showed that adults and infants can exploit an *aXb* structure to identify *a_b* dependencies, Mintz (2002, 2006) showed that the same structure could be exploited to (successfully) categorize the *X* elements. While the former type of learning arguably relies on ignoring the variable intervening *X* and focusing on the stable *a/b* elements, the latter requires focusing on the *X* elements and noticing their occurrence in the same context. Dependency-learning and category-formation, therefore, may be complementary processes that require attention to be distributed to different aspects of the input, either the string-peripheral or the string-internal items (respectively). It is possible that infants cannot direct their attention to these two types of regularities simultaneously, at least within a limited time of exposure such as the one in this experiment.

Soderstrom, White, Conwell & Morgan (2007) showed that 16-month-olds could track dependencies between functional morphemes over familiar words (i.e. discriminate ***These*** *chair***s**… from ****These*** *chair***_**…), or in passages with both familiar and unfamiliar intervening words (*These chairs…, These meeps…*) but not over unfamiliar/nonce words alone (***These*** *meep***s** vs. ****These*** *meep***_**…). Thus, if the target dependency was found only in contexts where it straddled a nonce word, infants' attention was captured by the novelty of the intervening nonce word and they failed to discriminate the correct and incorrect dependencies (see also de Diego-Balaguer, Martinez-Alvarez & Pons, 2016). A similar account could hold for the current experiment. Although infants may have become sensitive to the *a_b* dependencies at familiarization, this sensitivity may have been obscured at test when attention was directed towards the novel intervening *X'*, and away from the crucial *a_b* dependencies. Because previous work has suggested that sensitivity to NADs requires attention to the non-adjacent elements (Pacton & Perruchet, 2008), the current results of Experiment 1 may be explained not by a failure to learn the dependencies but by a failure to attend to them at test. Future research could address this confound by familiarizing infants with the novel *X'*s separately (i.e. not embedded in *aX'b* strings). If infants are familiar with the *X'*s, but the *aX'b* strings are still novel, then the generalization of *a_b* dependencies can be dissociated from the novelty of the individual *X'*s themselves.

4.2. Experiment 2

4.2.1 Introduction

The second experiment of this infant study sets out to replicate the adult findings in Chapter 2 (Experiment 2). There I showed that adult learners can successfully detect dependencies between prosodically non-salient units over intervening words that are prosodically more marked, as long as these dependencies are integrated into prosodically natural phrases. This was particularly relevant because NAD-learning has been hypothesized to serve the acquisition of morpho-syntactic dependencies in natural languages, which are instantiated between prosodically reduced elements – functors (which are often prosodic clitics, or affixes), spanning lexical words with lexical stress and higher prosodic prominence. In this experiment we aim to establish if 18-month-olds are equally capable of applying NAD-learning to contexts similar to natural languages, where the dependent elements are not prosodically prominent.

The theoretical premises of this study are the same as in Chapter 2 (see the Introduction of Chapter 2 for a more detailed discussion). Functional and lexical morphemes differ in their prosodic properties (Selkirk, 1996), and this distinction is marked by a variety of perceptual cues (Morgan, Shi & Allopenna, 1996; Shi, Morgan, Allopenna, 1998; Monaghan, Christiansen & Chater, 2007); these cues can be employed by infants from the moment of birth to distinguish between the two categories (Shi, Werker & Morgan, 1999). Although infants familiarize themselves with most of the functional morphemes of their language by the end of their first year of life (Shi, Cutler, Werker & Cruickshank, 2006; Shi & Lepage, 2008; Hallé, Durand & de Boysson-Bardies, 2008; Marquis & Shi, 2012), they also show a preference for listening to lexical over functional words (Shi & Werker, 2001, 2003), and an early inability to encode functors in full phonological detail (Stromqvist, Ragnarsdóttir & Richthoff, 2001; Shi, Marquis & Gauthier, 2006; Shi & Lepage, 2008) suggesting that their attention is captured by the more perceptually salient elements in their input.

The research question I address in this experiment is whether the specific perceptual cues to functional elements facilitate the detection of dependencies between them, or whether the lack of perceptual salience of these functors makes infants less likely to attend to the relationships between them. Just as in Chapter 2 I test the hypothesis that *Gestalt principles of similarity* are conducive to detecting NADs (Newport & Aslin, 2004; Creel, Newport & Aslin, 2004; Gebhart, Newport & Aslin, 2009), whereby dependencies are more easily learned between elements that perceptually similar to each other but distinct from the environment. I verify whether infants, like adults, can learn dependencies between both prosodically salient and prosodically reduced elements, or whether for infants, unlike adults, the perceptual salience of the dependent elements is crucial to detecting the dependencies.

4.2.2 Materials

The nonce words used in this experiment were identical to those in Experiment 1. The stimuli were created in the same way as in Experiment 2 of Chapter 2: both *a* and *b* tokens were recorded in carrier sentences in the position of functional morphemes preceding or following a noun stem, within a noun-phrase in the direct-object slot of a Subject-Verb-Object sentence. Thus, while *a* was recorded in the slot of the determiner preceding the noun (2), *b* was recorded as the suffix on the noun stem (3).

(2) *Ik zie __ zebra'tje.*

I see ___ little zebra.

e.g. *Ik zie* ***tep*** *zebratje.*

(3) *Ik zie het zebra__.*

I see the zebra__.

e.g. *Ik zie het zebra**lut**.*

The *X* stimuli for Experiment 2 were the same as the ones employed for Experiment1, and all the stimuli for Experiments 1 and 2 were recorded in the same session. Acoustic properties of the new *a* and *b* tokens, as analyzed using Praat 5303, are presented in Table A4.2 in the Appendix. The languages L1 and L2 were constructed in the same way as for Experiment 1, except the pauses between the words in an *aXb* string were not 250 ms but 100 ms. This experiment set out to emulate the New Functional Condition (100 ms) from Chapter 2. The test items were not novel *X*' stimuli from Experiment 1: instead of expecting infants to generalize to novel *aX'b* strings, in the test phase I used strings taken from the familiarization phase. The three *X* items from the familiarization that were also used at test were *domo, kasi* and *wadim*. As in Experiment 1, the 48 familiarization strings were played in random order with 750ms pauses in between. The familiarization phase lasted 1 minute and 40 seconds (shorter than in Experiment since the *a/b* tokens themselves were shorter in duration).

4.2.3 Participants

Infants were recruited in the same way as before, by written invitation to the parents who signed up in the UiL-OTS institute's database. A total of 40 infants were included (20 females, 20 males), with an average age of 18 months and 15 days (range: 18 months and 2 days – 19 months). All infants included had normal birthweight (2500-4500 grams), were not pre- or post-term (had a gestation period of 35-42 weeks), and had no known neurological, hearing or vision problems.

An additional 34 infants were tested but not included due to: low birthweight (*n* = 1), failure to retrieve information about birthweight and gestation period from parents (n = 3), fussiness, restlessness, crying or fatigue (n = 19), completing fewer than 2 valid consistent or 2 valid inconsistent trials (n = 2), parental interference or presence of distraction (n = 4) or technical error (n = 6). Again, drop-out due to fussiness was comparable to previous studies (19/74, 25.7%), and exclusion due to short looking times (completing fewer than 2 valid trials of each type) was very low (2/74, 2.7%).

From a total of 320 test trials of the 40 infants included (8 test trials per infant), 44 were excluded because they totaled a looking time shorter than 2 seconds. An additional 6 trials were excluded due to parental interference (n = 1), fussiness (n = 1), and experimenter error (n = 4), leaving 270 valid trials (84.4%).

Parents again completed the Dutch version of the MacArthur-Bates Communicative Development Inventory (CDI; Zink & Lejaegere, 2002), to establish receptive and productive vocabulary size at 18 months, and also offered informed consent of the participation of their infants in the experiment and of the use of the video recordings to analyze the data.

4.2.4. Procedure

The procedure was identical in all respects to that of Experiment 1.

4.2.5 Results

Table C4.2 shows the mean looking times (across trials and across infants) to test trials consistent or inconsistent with the language of exposure as well as the mean N-CDI scores (raw and percentiles), and the total looking time at familiarization, quantifying the amount of attention infants had during exposure to the language. As can be seen, infants again varied greatly, both in terms of language development (receptive and productive vocabulary) and in terms of their attention at familiarization. It is important to note that the looking time at familiarization in this experiment is shorter than Experiment 1 because the familiarization itself was shorter (1 minute and 40s instead of 2 minutes). In both Experiment 1 and Experiment 2 infants oriented to the blinking sidelights about 50% of the time at familiarization.

	Mean (Standard Deviation)	**Range**
LT Consistent	M = 10.622 s (SD = 5.882) Median = 7.72 s	3.41 – 26.63 s
LT Inconsistent	M = 10.676 s (SD = 5.401) Median = 6.88 s	4.05 – 25.84 s
Productive vocabulary	M Raw = 54.06 words (SD = 40.45) M Percentile = 54.09 (SD = 20.32)	10-158 words 15th – 90th perc.
Receptive vocabulary	Raw M = 229.52 words (SD = 106.02) Perc. M = 55. 88 (SD = 28.5)	72 – 473 words 10th – 99th perc.
LT at Familiarization	M = 50.131 s (SD = 16.60)	24.86 – 93.4 s

Table C4.2. Results of Experiment 2

I used a Linear Mixed Model with (a logarithmic transformation of) LT as dependent variable, Subject and as random intercept and Trial Type as fixed factor, and introduced the other factors (Language, Trial Number) one by one in the same way as before. The best model included Trial Number as covariate and Trial Type as fixed factor, and only yielded a significant effect of Trial Number (F (1, 232.96) = 10.127, p = .002) and no significant effect of Trial Type (p = .478). Thus looking times to Consistent and Inconsistent Trials did not significantly differ; however, children showed a decline of visual attention across the test phase again, with shortest looking times during the last (8^{th}) trial.

As in the first experiment, I also considered the possibility that infants may have been taking the first test trial as a cue for looking behavior across the subsequent trials (Gómez, Bootzin & Nadel, 2006). Therefore, I excluded the first trial and reanalyzed the data introducing First Trial (Consistent, Inconsistent) as a factor. The best model included First Trial, Trial Type and their interaction. There was a significant interaction of First Trial* Trial Type (F (1, 131.1) = 7,355, p = .007), and no significant main effect of First Trial (p = .716) or Trial Type (p = .237). Figure C4.1 illustrates the First Trial* Trial Type interaction: infants who heard a first test trial consistent with the familiarization showed a preference for consistent stimuli in the subsequent test trials, while infants who heard and inconsistent test trial first showed a subsequent preference for inconsistent stimuli.

It seems that, similar to Gómez, Bootzin & Nadel, (2006), infants developed a preference for the stimulus type they heard during the first test trial and maintained that preference throughout the following test trials. Out of the 20 infants who heard an inconsistent first test trial, 12 subsequently showed a preference for inconsistent strings; out of the 20 who heard a consistent test trial first, 16 showed a subsequent preference for consistent strings. Figure C4.1 shows the looking preferences for infants who heard a consistent vs. inconsistent test trial first:

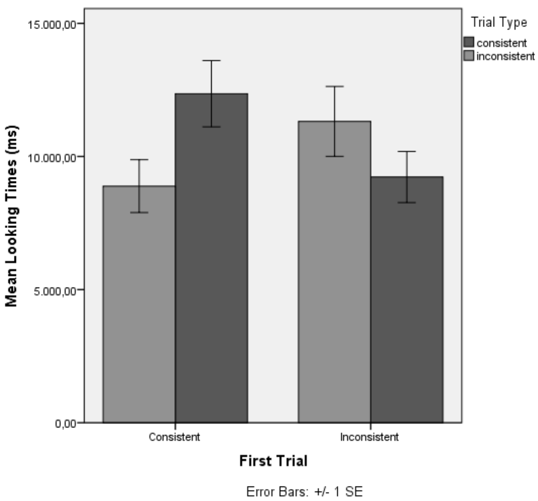

Figure C4.1 – Average looking times to trials consistent or inconsistent with familiarization (2-8) depending on the nature of the first trial (Consistent/Inconsistent); errors bars show the standard error of the mean (SE)

Given the significant First Trial by Trial Type interaction found, I split the data according to First Trial and analyzed each group separately with a Linear Mixed Model same as before. For the infants who heard a Consistent trial first, the best model included Trial Number and Trial Type, with both Trial Number (F (1, 95.297) = 7.55, p = .007) and Trial Type (F (1, 94.726) = 7.712, p = .007) being significant predictors. Thus, infants who heard a Consistent trial first showed significantly longer looking times to Consistent trials subsequently. For the infants who heard an Inconsistent trial first, the best model included Trial Type, which was not a significant predictor (p = .348) were significant. Thus, infants in this group did not show a significant preference of one type of stimuli over the other in trials 2-8, although Figure C4.1 suggests a non-significant preference for Inconsistent stimuli.

4.2.6 Discussion

In this experiment I asked whether infants can detect remote dependencies between prosodically distinct but reduced elements over prosodically more salient intervening material. Like in Experiment 1, I found no overall significant preference for one type of test item over another, but upon closer scrutiny the data revealed an interesting pattern.

Participants exposed to a first test trial that was consistent with the familiarization language maintained a significant preference for consistent trials throughout the remainder of the test phase. This in itself is evidence that infants could discriminate test trials based on the crucial difference under investigation: the specific a_i to b_i mapping in the a_iXb_i strings. On the other hand, participants exposed to a first test trial that was inconsistent with the language of exposure maintained a non-significant preference for inconsistent trials. It seems that infants took their cue from the initial test trial but they did so more robustly when this initial trial was consistent with what they had been heard at familiarization.

The finding that infants use the first test trial to guide their preference for the rest of the test trials is not new. Gómez, Bootzin & Nadel (2006) found a similar effect in a NAD-learning experiment where infants napped between the familiarization phase and the test. They familiarized 15-month-olds with an *aXb* language for 15 minutes in their home and then tested them on recognition of the *a_b* dependencies in familiar *aXb* strings 4 hours later. Infants who did not nap (or napped for under 30 minutes) between familiarization and test exhibited a significant preference for the correct dependencies, as predicted (Gómez & Maye, 2005). However, infants who napped (for more than 30 minutes) showed a distinct pattern. They did not show an overall preference for one type of stimulus or the other, instead they showed a significant preference for the type of stimuli that they heard in the first test trial. Furthermore, infants who napped but were exposed to an *aXb* language with low variability of the *X* elements (3, which was shown not to support NAD-learning cf. Gómez, 2002, Gómez & Maye, 2005), showed neither pattern, suggesting that simply hearing the first test trial did not induce a preference in the subsequent test trials, but that retention from the initial familiarization was crucial to learning.

The authors concluded that when infants learned at familiarization and napped subsequently, sleep erased the item-specific memory traces of the $a_i\_b_i$ mappings (e.g. *tep* goes with *lut* and *sot* goes with *jik*). Gómez & Edgin (2015) attribute this to the fact that sleep before 18 months does not seem to support (item-specific) memory consolidation (which develops between 18-24 months), but rather seems to facilitate generalization (abstracting the knowledge that a dependency exists, without specifying the dependent elements). Thus the abstract representation of a dependency between the first and last word in an *aXb* string could have been retained and activated in the first test trial, where infants could use it to identify the specific $a_i\_b_i$ mappings from just a short amount of exposure.

There are crucial differences between the Gómez et al. (2006) study and Experiment 2. Experiment 2 presented 18-month-old (not 15-month-old) infants with a shorter exposure (only 2 instead of 15 minutes) to a language with perceptually non-salient *a/b* tokens, had a much shorter break between familiarization and test (a contingency training phase of no more than 30s, but where infants continued to receive stimulation in the form of tones) and did not involve sleep. Importantly, in Experiment 2 the effect of the first trial on preference in the subsequent trials was only robust for first trials consistent with familiarization.

Nevertheless, I propose that Gómez et al.'s (2006) interpretation of their results could be extended to the current findings as well. Infants in this experiment were presented for a short period of time to a language where the target words instantiating the regularity to be learned were prosodically 'reduced'. As pointed out before, infants between 6 and 12 months of age have difficulties encoding functional-sounding words or morphemes (prosodic clitics; Stromqvist, Ragnarsdóttir & Richthoff, 2001; Shi, Marquis & Gauthier, 2006; Shi & Lepage, 2008). Although after their first birthday infants become sensitive to the phonological detail of many of the functors in their language (presumably due to prolonged exposure), it is not unlikely that they retain a difficulty to phonologically encode prosodically reduced elements from only brief exposure to them. It is furthermore not surprising that the fragile phonological representations obtained from this brief exposure should deteriorate over an intervening contingency training phase, where infants are required to attend to a different type of stimuli.

Therefore, even without a sleep episode intervening between familiarization and test, infants in this experiment may have, at least partially, lost the item-specific representations of $a_i\_b_i$ dependencies. Subsequently, they used the first test trial to recover these representations and re-establish the correct dependencies. This interpretation crucially rests on the assumption that 18-month-olds commit fragile or incomplete representations of prosodically reduced (functor-like) elements to memory, and that these memory traces are easily deteriorated by subsequent exposure to something else, in this case, to a pure tone and the association between looking behavior and sound/visual stimulus. Further research is necessary to confirm this assumption.

The second question is why infants in Experiment 2 did not show a *significant* preference for *inconsistent* stimuli (with respect to familiarization) when the first trial exposed infants to such stimuli. The asymmetry in these findings suggests that the memory representations for the specific $a_i\_b_i$ dependencies were only partially deteriorated during the intervening contingency training phase. It may be that infants retained a (phonologically) underspecified memory trace for each of the $a_i\_b_i$ dependencies. Thus, when the first test trial was consistent with familiarization, it confirmed and strengthened these earlier representations, filling in potentially missing features. However, when the first test trial was not consistent with familiarization, the infants may have been confused. Because the new input did not match their pre-existing representations, they could not (always) take their cue from the first test trial. However, because these pre-existing representations may have been fragile and incomplete, they may not have sufficed to guide the infants' preference. If this is the case, then the prediction is that infants' verbal memory (their ability to memorize words from limited exposure) may be inversely correlated with their ability to use test trial 1 to guide their preference during test trials 2-8. Unfortunately, because I have no such measure for the infants in this experiment, this predicted correlation should be investigated in future research.[11]

[11] Note also, that even though the preference for inconsistent stimuli was not significant in the Inconsistent First Trial group (perhaps due to a lack of power), it was still in the right direction, showing longer looking times during inconsistent trials. Perhaps a longer intervening phase between familiarization and test could accentuate the 'deletion' of item-specific memory traces and aid learners in developing a new preference during the first test trial.

To conclude, Experiment 2 showed some evidence that 18-month-old infants may be able to detect dependencies between prosodically 'reduced' (and over prosodically salient) elements in an artificial grammar. A statistically significant interaction between Trial Type and First Trial revealed that infants showed different patterns of preference depending on the material they were exposed to in the first test trial. Subsequent analyses revealed that infants showed a significant preference for consistent over inconsistent test trials when the first test trial was also consistent with the familiarized language. However, infants' preference for inconsistent stimuli when the first test trial was inconsistent itself was shown not to be statistically significant. I proposed an account for these findings based on Gómez et al. (2006), where infants acquired fragile/incomplete item-specific representations of the $a_i\_b_i$ dependencies, and required the first test trial to consolidate these representations in order to robustly discriminate them from inconsistent $a_i\_b_j$ dependencies.

4.3 General Discussion

The ability to detect remote dependencies in spoken input has been attested with infants as young as 15 months old (Gómez & Maye, 2005), and has been hypothesized to contribute to the acquisition (starting from around 18 months, Santelmann & Juszcyk, 1998) of morpho-syntactic dependencies in natural languages. NAD-learning has been shown to correlate with processing of remote syntactic relationships in adults (Misyak, Christiansen & Tomblin, 2010; Misyak & Christiansen, 2012) and appears to be less robust, or delayed, in 18-month-olds at familial risk for dyslexia (Kerkhoff et al., 2013). The current study set out to find new evidence that NAD-learning is indeed a powerful learning mechanism that may help infants acquire stable rule-like representations of morpho-syntactic relationships in natural language.

The two research questions tackled in this study where: (i) Can infants generalize the NADs they have learned to novel contexts? and (ii) Can they detect NADs between prosodically non-salient dependent elements? Because adults proved capable of both tasks, and because infants exhibit an advantage in language learning compared to adults, I predicted that infants would show significant learning in both cases. However, results did not always confirm these expectations.

While Experiment 1 showed no evidence of generalization of NADs to novel contexts, Experiment 2 showed some evidence that NADs between prosodically 'reduced' items can be detected in the input. Infants exposed to an

a_iXb_i language with functor-like *a/b* tokens employed the first test trial in a subsequent test phase to guide their discrimination of correct a_iXb_i vs. incorrect a_iXb_j strings. Infants who heard an initial test trial consistent with the familiarization language subsequently oriented significantly longer to consistent stimuli, indicating that they could discriminate between consistent and inconsistent strings. Infants who heard an inconsistent first trial also oriented longer to inconsistent stimuli, but this preference was not statistically significant, suggesting that the inconsistency between the familiarization and the first test trial may have confused (some) infants. Overall I concluded that infants could detect NADs between prosodically reduced elements, although the memory traces of these specific $a_i\_b_i$ dependencies might deteriorate over time without sufficient repetition. In what follows I propose that the role of *attention*, which has been recently investigated in NAD-learning, may explain some of the trends in our data.

Pacton & Perruchet (2008) pointed out the importance of attention to NAD-learning. In their tasks (adult) learners identified remote dependencies in continuous strings of digits only when they were specifically instructed to mentally compute arithmetic operation between non-adjacent digits. Endress et 2005) also argued that NADs are more easily learned when the dependent elements occur at string edges – positions that are marked by pauses, and are therefore more easily attended to. In Chapter 2 we saw that, although adults can learn dependencies between both highly salient and highly reduced elements, they show a greater learning effect with salient stimuli, suggesting that perceptual salience might facilitate NAD-learning (perhaps in directing the learner's attention to the target elements).

De Diego-Balaguer, Martinez-Alvarez & Pons (2016) proposed that all these cues (explicit instruction, string-peripheral position or prosodic cues) pertain to exogenous attention which can direct the learner towards remote dependencies when there is no other bias to attend to other aspects of the input. Exogenous attention is present from birth and influences early language acquisition by making infants tune in to salient aspects of the input, such as prosody. In our experiments factors modulating exogenous attention may have prompted infants to focus on the intervening element *X* rather than the dependent elements *a_b*. In Experiment 1, the *Xs* were novel at test (while the *a/b*s were familiar), and therefore may have engaged infants' attention in categorizing them as similar with the previous *X*s. In Experiment 2, the *X*s were simply more prosodically salient than the *a/b* items.

A different type of attention, endogenous attention, may support NAD-learning in a different way according to De Diego-Balaguer et al. (2016). Endogenous attention is the attention learners pay to certain aspects of the input which they know, from previous experience, can be relevant. De Diego-Balaguer et al. (2016) suggest that endogenous attention develops later than exogenous attention, around the same time that infants begin to exhibit the ability to detect NADs. Adult participants in Experiment 2 of Chapter 2 may have learned dependencies between prosodically reduced elements because their prior linguistic experience told them that these elements are highly informative to syntactic structure, and can often occur in remote dependencies. Thus, their attention was already endogenously directed to tracking these elements, and (in particular) any non-adjacent relationships they may have entertained.

The infants in Experiment 2 were 18-month-olds, an age at which they are also beginning to familiarize themselves with morpho-syntactic dependencies in their own native language (Santelmann & Jusczyk, 1998). Their sensitivity to the dependencies between non-salient *a/b* items could also be explained by endogenous attention to functor-like elements. Whether or not infants can detect dependencies between prosodically non-salient items without relying on their previous linguistic experience is a question that remains open for further investigation. If younger infants (15 months) show learning with the same type of stimuli as in Experiment 2, this should support Newport & Aslin's (2004) hypothesis that, in the absence of any biases based on prior linguistic experience (as 15-month-olds show no sensitivity to morpho-syntactic dependencies in their native language, cf. Santelman & Jusczyk, 1998), learners are guided by Gestalt principles of perception in their acquisition of NADs.

To conclude, I found no evidence that 18-month-olds were capable of detecting remote dependencies *and* subsequently recognizing them in novel contexts. However, there was evidence that remote dependencies, even between perceptually reduced elements, were recognizable in familiar strings. The results of the two Experiments in this chapter may be accounted for if we consider the role of attention in NAD-learning, but future research is called upon to identify the different factors that may have influenced infant behavior in these experiments.

Chapter 5

Discussion and Conclusions

In the studies presented in this dissertation I have set out to assess the potential role that NAD-learning (the human ability to learn dependencies between non-adjacent elements in the input) could play in the acquisition of morpho-syntactic dependencies in natural languages. The aim was to understand whether the particular properties of morpho-syntactic dependencies and the timeline of their acquisition would support a distributional learning mechanism that can detect co-occurrence patterns between non-adjacent elements in the input, and, conversely, whether the properties of the learning mechanism would make it a viable tool for language acquisition.

Two main proposals were investigated in this series of studies. The first started from the observation that morpho-syntactic dependencies are instantiated between functional morphemes that are marked by a constellation of perceptual cues, and which are less prosodically salient in the input than lexical morphemes. If NAD-learning can be successfully recruited by infants to glean the morpho-syntactic dependencies in their native languages, then it should support the detection of dependencies even between perceptually 'reduced' elements. Therefore I investigated the role of prosodic cues in NAD-learning for both adult and infant learners. In section 5.1 I discuss these findings and their implications in detail.

The second proposal arose from the observation that, behaviourally, infants show sensitivity to individual functional morphemes at a younger age than they show sensitivity to the dependencies between them.[12] It seems, therefore, that infants may build their representations of morpho-syntactic dependencies on pre-existing representations of the functional morphemes that enter such dependencies. I asked whether the NAD-learning mechanism is also built in such a way that prior familiarity with individual elements supports subsequent detection of the dependencies between those elements. In section 5.2 I discuss the findings from Chapter 3, and propose future research that should shed light on this matter.

Together, these two lines of research had a dual purpose: firstly, to clarify how NAD-learning could potentially subserve language acquisition. In section 5.3 I discuss how the findings in Chapters 2-4 shed light on the potential role of NAD-

[12] Although some neurophysiological evidence exists that infants can develop a sensitivity to NADs at a much younger age, cf. Friederici et al (2011)

learning in the acquisition of morpho-syntactic dependencies. A second aim of these studies was to gain a better understanding of the NAD-learning mechanism, its properties and the constraints that limit its scope. Section 5.4 discusses what my findings reveal about the workings of the NAD-learning mechanism. Finally, the three studies in this dissertation have also revealed the importance of carefully choosing one's methodology. In the final section 5.5 I conclude the dissertation by discussing some fruitful methodologies and research directions that could bring us closer to understanding how humans learn dependencies between non-adjacent elements in linguistic input, and how this ability could be involved in language acquisition.

5.1 The Role of Prosody in NAD-Learning

In Chapter 2 I investigated the role of perceptual cues in the detection of NADs, and linked it to the perceptual cues that mark the distinction between functional and lexical morphemes in natural languages. Newport and Aslin argued that non-adjacent dependencies may be guided by Gestalt principles of perception, which allow perceptually similar elements to be 'grouped together' and processed as more strongly related than the linear distance between them would indicate. This would prompt the distributional learning mechanism to compute co-occurrence statistics preferentially over elements that are similar to each other (e.g. functional-like but distinct from the intervening lexical-like material). To demonstrate the claim that Gestalt principles of perception facilitate NAD-learning, one would have to demonstrate: (i) that NAD-learning is easier when the dependent elements are distinct in their environment than when they are perceptually similar to the intervening material; (ii) that NAD-learning is easier when dependent elements are similar to each other than when they are perceptually distinct. In Chapter 2 I focused mainly on (i).

In a simple *aXb* language I systematically varied the perceptual properties of the *a/b* elements while keeping the *X*s constant. I showed that when *a/b* where perceptually distinct from the intervening *X*, the *a_b* dependencies could be detected (provided the contour of the *aXb* strings was prosodically natural). However when the *a/b* and *X* elements were all perceptually similar, there was no evidence of learning. In a final experiment I also briefly addressed the claim in (ii): participants did not show significant above-chance learning in a condition where the *a/b* elements were perceptually distinct in their environment but also distinct from each other. Future research must elucidate whether the latter result is truly due to the difficulty of detecting dependencies between dissimilar elements, or to the specific way that the stimuli in this experiment were constructed. Various configurations in which the *a/b* are perceptually dissimilar should be tested (e.g. not just Emphasized_Lexical_Functional, but also Functional_Lexical_Emphasized, Functional_Emphasized_Lexical, etc.), and compared to conditions with minimally different languages in which *a/b* are perceptually similar.

A further finding of this study was that Gestalt principles of perception only applied in specific circumstances. That is, while participants could learn dependencies between perceptually distinct *a/b*s, either more or less salient than the intervening *X* elements, this learning was only possible in conditions with natural prosody (where perceptually 'reduced' *a/b* tokens were separated from *X* by short 100 ms pauses, and perceptually salient *a/b*s by long 250 ms pauses). Participants thus seemed to rely on unequivocal cues to the prosodic status of the dependent *a/b* elements. I interpreted these findings as indicating that participants were identifying items in the *aXb* strings simply as prosodically 'marked' or 'unmarked', and were subsequently using Gestalt principles to group together elements with the same prosodic properties.

Furthermore, I showed that given a natural prosodic contour to the *aXb* phrases, participants could detect dependencies either between highly perceptually salient *a/b* elements over less salient *X* elements, or between less salient (functional-like) *a/b* elements over more salient (lexical-like) intervening *X*s. Because the stimuli were specifically constructed to emulate the properties of functional and lexical elements in (a) natural language, this adult data directly supported the hypothesis that the specific properties that mark functional morphemes in natural languages facilitate the detection of remote dependencies between them.

Importantly, however, while learning was possible with both highly perceptually salient and highly non-salient dependent elements, adults showed a slightly larger learning effect when they were learning dependencies between highly salient elements. Therefore, while Gestalt principles of perception can guide NAD-learning, there might be an additional role for perceptual salience. Because there was a significant learning effect with both highly salient and non-salient *a/b* dependencies, the current study is not able to directly assess the role of salience in learning *a_b* dependencies. Future research should replicate the two successful learning conditions (the Emphasized Condition with 250 ms pauses and the New Functional Condition with 100 ms pauses, see Chapter 2) with a larger sample size for each condition, and determine whether salient dependencies are significantly easier to detect.

The result with 18-month-olds (see Chapter 4, Experiment 2) suggested that infants too are sensitive to dependencies between functional-like (over lexical-like) elements, but that here too perceptual salience might be relevant. Infants showed a significant difference in their looking preference depending on the nature of the first test trial. When exposed to a first test trial consistent with the familiarization they showed a significant preference (across test trials 2-8) for the dependencies at familiarization. When exposed to an inconsistent first test trial they showed a subsequent preference for dependencies inconsistent with familiarization (though the latter effect was not statistically significant). Previous studies have observed similar patterns in which infants used the first test trial to guide their preference in the subsequent trials (Gómez, Bootzin & Nadel, 2006; Hupbach, Gómez, Bootzin & Nadel, 2009). It has been claimed that infants retain an abstract representation of dependencies from familiarization, but fail to memorize the item-specific one-to-one mappings between dependent elements. A similar interpretation is possible for the results of the current study. Infants may have detected dependencies at familiarization. However, because the dependent elements were not perceptually salient, and because familiarization was separated from the test phase by a (brief) contingency training phase, the memory traces for dependencies between specific items may have quickly decayed during the contingency training phase. At test, infants may have used the first test trial to re-familiarize themselves with the specific dependencies.

One caveat for both the adult and infant experiments is the confounding factor of prior linguistic experience. At the age of 18 months, infants are already becoming familiar with some of the morpho-syntactic dependencies in their native language. Their (emerging) linguistic experience may therefore inform them that highly frequent but perceptually non-salient elements are important markers of grammatical structure and may often form remote dependencies. For adults too the results of Experiment 2, Chapter 2 may be explained if we assume that the participants were identifying the (Dutch-functor-sounding) *a/b* elements with functional morphemes in their native language (Dutch), and this facilitated the inference that there might be dependencies between them.

This account could, in principle, explain the findings with infants and part of the findings with adults, but additional assumptions would be necessary in order to explain why adults learned dependencies between highly acoustically emphasized *a/b*s, which are less likely to occur frequently in natural languages, especially in the participants' native language Dutch. So far, the Gestalt principles of perception account seems more parsimonious, although future research should strive to disentangle the learning effect found in this study from any possible influence of linguistic experience. It should be pointed out that this is a frequent and seldom signaled problem with artificial grammar learning, especially with adults but also with infants. Whenever subjects are presented with artificially created stimuli whose structure emulates some property of natural languages, there is a chance that their sensitivity to that structure might be (partly) prompted by their experience with it in their native language. In the present case, familiarity with remote dependencies, although behaviorally evidenced around 18 months of age, has been shown (using neurophysiological methods of investigation) to potentially arise as early as 4 months (Friederici et al., 2011). Therefore, only in testing the effect of perceptual cues in this very young age group, using EEG or other neurophysiological methods, could we hope to truly test the effect of Gestalt principles of perception in the absence of any biasing influence of prior linguistic experience.

5.2 The Role of Incremental Learning in NAD-Learning

Chapter 3 approached the role of NAD-learning in language acquisition from a different perspective. I started from the observation that behavioural studies show an early sensitivity to functional elements that precedes the sensitivity to morpho-syntactic dependencies between them, and that therefore it is possible that dependencies between functional elements are learned at a point in linguistic development when the individual elements themselves are already familiar to the learner. The study in Chapter 3 was aimed at understanding whether this incremental learning trajectory (from individual elements to patterns between them) facilitated NAD-learning. Two alternative hypotheses were proposed. According to the Less-Is-More, or Starting Small Hypothesis (Newport, 1990; Elman, 1993), learning complex patterns between elements should be facilitated by prior familiarity with the individual elements. According to the Starting Big Hypothesis, on the other hand (Arnon, 2009; Arnon & Ramscar, 2012) forming associations between elements is inhibited by prior familiarity with the elements in isolation.

Chapter 3, thus, addressed a question never directly tackled before: whether prior familiarity with individual elements (*a,b*) facilitates the subsequent detection of the (*a_b*) remote dependency between them. This line of research would give us a clearer understanding of the workings of the NAD-learning mechanism, as well as making specific predictions for language acquisition. If dependency-learning is, for instance, inhibited by prior familiarity with the dependent elements, then NAD-learning could only be a viable mechanism for language acquisition so long as early learners start tracking functional morphemes at the same time as developing their sensitivity to the dependencies between them.

The first series of experiments seemed to confirm that this was, in fact, the case. Participants who were pre-familiarized with *aY/Zb* phrases (containing the target *a/b*s) and subsequently familiarized with the *aXb* language showed declining performance in learning the language with longer exposure to the pre-familiarization. In a second experiment, instead of increasing the *amount* of exposure I increased the *salience* of the target words at pre-familiarization, familiarization and test. Comparisons with previous findings showed that with less salient (functor-like) *a/b* elements, the presence or absence of a one-minute pre-familiarization did not make a difference for learning. On the other hand, with highly salient (emphasized) *a/b* words, the introduction of a one-minute pre-familiarization significantly inhibited learning.

The results of the first two studies suggested a negative effect of prior familiarity. In the conditions where participants were pre-familiarized more extensively, or more saliently with the target words of the language, they failed to subsequently track dependencies between these words at familiarization. These results were consistent with the Starting Big Hypothesis that any association between *a* and *b* was blocked if the two were learned separately beforehand. However, there was also the possibility that the very existence of a pre-familiarization, long or salient enough to engage participants in detecting adjacent combinatorial patterns, would subsequently make learners less engaged in detecting non-adjacent patterns at familiarization. To exclude this possibility, I ran a third experiment where the pre-familiarization consisted of a simple list-wise, one-minute presentation of non-salient *a/b* words. The results of this experiment confirmed that the very occurrence of a pre-familiarization phase inhibited learning: participants exposed either to the target *a/b* words or nonce monosyllabic foils were equally unable to show above-chance learning.

In the fourth and final experiment I sought to explain why the very presence of a pre-familiarization phase might have a negative effect on learning, and whether this confounding effect could be obviated. Previous research showed that adults can learn distinct artificial languages sequentially (Gebhart, Aslin & Newport, 2009) only if the transition between the distinct languages is marked by sufficiently salient cues. Therefore, participants in my experiments might have only tuned in to the strings they heard at pre-familiarization, gradually lost their attention to the input and failed to notice the transition between pre-familiarization and familiarization. I found that participants who received explicit instructions about the transition between learning phases performed better overall than those who were not explicitly instructed[13].

[13] Note that the necessity for explicit instruction is linked to the sequential presentation, in the same learning session, of two types of input – a setup which previous literature (Gebhart, Aslin & Newport, 2009) has attributed to a decline of attentional resources as learners become familiarized with the first type of input. Arguably, the same need for an explicit separation of learning phases should not arise during language acquisition, where

I also tested the effect of prior learning by comparing an experimental condition (where participants were pre-familiarized with the target *a/b* words) with an analogous control condition where the target words were replaced with 6 other monosyllabic ones. Unfortunately, a statistically significant difference between the two pre-familiarization conditions (control and experimental) was not observed. There seemed to be a slight numeric advantage to the experimental condition (a mean accuracy rate of 62.7% compared to 57% in the control), suggesting that, contrary to the previous findings, prior familiarity with words might in fact boost subsequent learning of dependencies between them. In order to consolidate these findings, I am currently investigating this effect with even more explicit separation of the pre-familiarization and familiarization phase, with a more informative pre-familiarization and a larger number of participants per condition.

Although the experiments of this chapter did not yield a clear answer to the main research question, they provide important cues as to how future research should address this question. Previous research looking into the role of incremental exposure (see references in Chapter 3) seldom explicitly considered the possibility that the transition from one learning phase to the next may inhibit learning in the latter. Arnon & Ramscar (2012), for instance, compared adjacent dependency-learning between a group first exposed to the words and subsequently to the strings, and a group first exposed to the strings and subsequently to the words. If participants showed a primacy effect, as in the current study and previous studies, then it is no surprise that participants first exposed to the strings were better at learning the regularities within them.

The findings in my Experiment 4 therefore inform future research by revealing the importance of the timeline of an artificial grammar learning experiment. Participants exposed to multiple exposure phases one after the other may need strong cues to recover their attention to the input at each transition between phases. Even with such cues, their attention may decline over time. Participants in the first experiment were explicitly instructed about the transition between the two learning phases and still failed to show learning in the conditions with longer pre-familiarization. Even with transition cues, lengthy exposure prior to the target language may introduce fatigue and diminish the ability to learn at familiarization. The nature of the pre-familiarization may also pre-dispose participants to attune to the specific regularities present at this stage and to subsequently ignore all other types of regularities. Finally, if some but not all

infants learn over multiple sessions, as their working memory capacities expand. A better model of this would be to have the pre-familiarization and familiarization in separate learning sessions.

words of the language (the *a/b* words but not the *X*s) are presented at pre-familiarization, then during familiarization participants might be led to attend to the novel words preferentially, and fail to notice regularities between the familiar words.

In light of all these considerations, in Chapter 3 I proposed that a more reliable procedure would separate the pre-familiarization and familiarization phases, by (extensively) pre-familiarizing the participants to all the words of the *aXb* language on one day and then familiarizing them to the language itself on the next day. Such a set-up would not only help eliminate confounding factors like fatigue, loss of attention, preference for novel words, etc. but would also make the experiment ecologically more valid, since memory traces for the words would be encoded in long-term memory (if they survive to the second day of testing) in the same way that infants may encode functional morphemes in long-term memory before they begin detecting remote dependencies between them.

Furthermore, an investigation of individual differences in learning might prove fruitful. As pointed out in Chapter 3, it is possible that while some learners are inclined to 'start small' (by building the representations for regularities in the input on pre-existing representations of the units between which these regularities obtain) and thus might benefit from incremental input, others may be inclined to 'start big' and might thus not be aided by incremental input. Differences between Analytic and Holistic learners have been identified in the study of Cognitive Styles (Riding, 1991; Riding & Rayner, 1998; Peterson, Deary & Austin, 2003), and it has been shown that cognitive style affects the way that learners benefit from staged input (Riding & Sadler-Smith, 1992). Thus, including individual measures of cognitive style (such as, for instance, Riding's 1991 Cognitive Styles Analysis) might be helpful in identifying whether NAD-learning proceeds differently for different types of learners, or whether Starting Big or Small universally supports learning of remote dependencies in all learners.

5.3 Generalizing NADs to Novel Contexts

Aside from the individual research questions of my two lines of investigation with adults, the studies presented above also show consistent evidence that adults are capable not just of retaining specific *a_b* dependencies, but also of tracking these dependencies in novel contexts, with intervening material they had not heard before. In all of the adult experiments presented here I tested participants on their ability to identify the target *a_b* dependencies in novel *aX'b* strings. When learners acquire specific morpho-syntactic dependencies in natural languages (e.g. *The princess* ***is*** *gently kiss**ing** the frog*), full mastery of these dependencies means not just recognizing them in familiar contexts (e.g. with familiar intervening verb stems or adverbs) but being able to generalize the rule to completely novel ones (e.g. *The slithy tove* ***is*** *gyr**ing** and gimbl**ing** in the wabe*). Otherwise learners may be simply memorizing chunks (…*is gently kissing…*) entirely, without identifying the dependency on its own or perceiving it as a fully productive rule.

Previous research has suggested that infants and adults may indeed be generalizing when learning NADs. Gómez (2002) showed that both adults and infants learned *a_b* dependencies in *aXb* strings only when the intervening *X* element was highly variable (and therefore infrequent). If her participants had been simply recalling fully memorized *aXb* strings, then more frequent and less variable *X* elements (and implicitly more frequent and less variable *aXb* strings) should have facilitated recall. Instead, learners seemed to exploit the variability of the *X* element in order to attend to the more invariant *a* and *b* elements and become sensitive to the regularities between them. The high variability of the *X* elements likely prompted a more abstract categorical representation for the intervening material, whereas the high frequency of the *a_b* dependencies supported a more item-specific representation of the one-to-one mapping between *a* and *b* words.

Therefore it comes as no surprise that adults in our experiments showed evidence of recognizing the relevant dependencies even in novel contexts with an unfamiliar intervening *X*. What is surprising, however, is that 18-month-old infants showed no evidence of doing the same. Experiment 1 of Chapter 4 tested 18-month-olds' ability to track remote dependencies and subsequently recognize them in novel *aX'b* strings. The experiment largely emulated previous studies (Gómez 2002, Gómez & Maye, 2005; Kerkhoff et al., 2013), making slight changes to the way the stimuli were recorded and introducing a brief contingency training phase to familiarize infants with the procedure prior to the test phase. Failure to obtain a reliable effect of learning could be due to these methodological changes, although it is important to note that in Experiment 2 of Chapter 4 I used the same procedure and obtained a preference across test trials 2-8 for stimuli consistent with those of test trial 1, which was not observed in Experiment 2. Without a control condition using an identical procedure to test a similar group of 18-month-olds on NAD-learning (but NOT generalization), it is difficult to determine whether the failure of Experiment 1 was due to the methodological changes (in which case a control condition would not have yielded a significant learning effect) or to actual difficulties in generalizing the familiarized dependencies to novel strings (in which case a control condition without generalization would have shown a significant improvement in learning).

If further testing using control conditions does reveal a difficulty in generalizing *a_b* dependencies at test, it may be easy to account for such a finding.

Soderstrom et al. (2007) showed that 16-month-olds found it easier to track morpho-syntactic dependencies over familiar (***these*** *chair**s***) rather than unfamiliar/nonce noun stems (***these*** *meep**s***). This may be because, in the latter case, the attention of the infants was captured by the novelty of the nonce noun stem and drawn away from the grammaticality of the dependency (de Diego-Balaguer, Martinez-Alvarez & Pons, 2016). Thus, infants may well retain the representation of dependencies as generalizable rules, but may not be able to always *apply* this knowledge, due to limitations of their attentional resources. A less challenging task (and one perhaps more similar to natural language acquisition) would require generalizing the *a_b* dependency to an *X'* which has never been used as an intervening element in an *aX'b* string, but is otherwise familiar to the infant from other contexts. This would imply that the *X'* is not completely novel, but the *aX'b* configuration is.

5.4 NAD-Learning as a Mechanism for Language Acquisition

One of the main research questions that prompted the research presented above was: can NAD-learning be viable as a mechanism for the acquisition of morpho-syntactic dependencies in natural languages? I started from the premise that if NAD-learning contributes to the acquisition of morpho-syntactic dependencies in a significant way, then (some of) the properties of these dependencies and the timeline of their acquisition are likely to serve as facilitative cues to NAD-learning.

In the first line of research, I obtained evidence that both infants and adults could detect dependencies between prosodically 'reduced' elements over a prosodically marked intervening element. This is particularly relevant given that in many languages functional morphemes are often prosodic clitics (Selkirk, 1996), perceptually 'reduced' compared to lexical morphemes (Shi, Morgan, Allopena, 1998; Monaghan, Christiansen & Chater, 2007). Thus NAD-learning can be successfully applied to the type of dependencies that are present in natural languages, despite the fact that these dependencies are instantiated between perceptually non-salient units, and lack of perceptual salience may make words harder to encode (Shi & Lepage, 2008; Stromqvist, Ragnarsdóttir & Richthoff, 2001).

Furthermore, the adult study in Chapter 2 showed that learners may be guided by Gestalt principles of perception: it is in itself remarkable that general principles of perception (shown to apply to non-linguistic input as well) can exploit the properties of natural languages (i.e. the lexical/functional distinction) to guide the NAD-learning mechanism towards the relevant morpho-syntactic relationships, at least in the cases where these properties/cues exist. It is also interesting that these Gestalt principles seem to apply not at a purely acoustic level, but at the level of prosodic representations. I believe a NAD-learning mechanism that can exploit prosodic cues is more efficient than one that can only exploit acoustic cues. As we saw from the stimuli created for Experiment 2 Chapter 2, functor-like elements can have slightly different acoustic properties depending on the context they are embedded in (e.g. the beginning or end of a prosodic phrase), but learners seem to abstract away from these differences and process the elements as similar on a deeper level (prosody).

It seems, therefore, that being guided by Gestalt principles of perception, and being able to apply these principles to the prosodic representations of units in the input renders NAD-learning a potentially effective mechanism for language acquisition. Nevertheless, it must be pointed out that morpho-syntactic dependencies in natural languages are not always instantiated between prosodically similar elements. As mentioned in Chapter 2, languages that assign prominence to the final syllable of a word may display dependencies between unstressed functional words and stressed functional suffixes, raising the question of whether Gestalt perception does not inhibit the acquisition of these types of dependencies. In Experiment 3 Chapter 2 I found no evidence of learning *a_b* dependencies between a prominent *a* and a non-prominent *b*; however, the learners in my experiment were Dutch natives, a language where many morpho-syntactic dependencies are indeed instantiated between prosodically 'weak' functional morphemes. Future research is necessary to determine whether all learners, irrespective of their native language, would apply Gestalt principles of perception to NAD-learning, or whether Spanish learners for instance would show a higher sensitivity than Dutch natives to dependencies between dissimilar dependent elements. It also remains to be seen whether Gestalt principles could inhibit the detection of certain types of dependencies in natural languages, or whether such dependencies can still be learned (perhaps with the help of other cues highlighting the dependent elements).

It is important to consider that Gestalt principles of similarity might apply not only at the level of auditory perception, but could facilitate NAD-learning between any elements with similar properties that are distinct in their environment. Gómez (2002), for instance, showed that *aXb* dependencies are more easily detected when the *a/b* elements are highly frequent and invariable, whereas the *X*s are highly variable and infrequent. Onnis et al. (2005) also showed that with a completely invariable *X* and three different *a_b* dependencies spanning it, the dependencies were still learnable. Gestalt principles, therefore, could perhaps also apply when analyzing distributional properties of the input: elements with similar frequencies may be grouped together and processed on a separate level from other dissimilar elements.

One particularly interesting question is whether Gestalt principles of similarity could also capitalize on categorical information to facilitate NAD-learning. If, as discussed in Chapter 3, functional morphemes are acquired and categorized as functional (as opposed to lexical) before the age where dependencies between them are learned, then this cue differentiating functors from the intervening lexical material could support dependency-learning.

In three out of four experiments in Chapter 3 I exposed participants to a pre-familiarization phase where learners could segment the *a/b* elements from two-word phrases, as well as observe that they combine productively with a variety of bisyllabic words, and that while the *a*-elements were always string-initial, the *b*-elements were always string-final. Although the findings of Chapter 3 were not conclusive, they open up an avenue for research that could prove fruitful and interesting. If pre-existing representations for *a/b* elements can facilitate the detection of *a_b* dependencies, then learning individual functors may bootstrap (NAD-)learning of morpho-syntactic dependencies. As in the case of my first study, this would again imply that the lexical/functional distinction in natural languages is particularly suited to cue NAD-learning.

Finally, full mastery of a morpho-syntactic rule, I have argued, implies the ability to apply it productively, in novel as well as familiar contexts. In Chapter 4 I failed to show that 18-month-olds were capable not only of detecting remote dependencies but also of recognizing them in novel contexts. However, a similar pattern to the current result was observed by Soderstrom et al. (2007), who showed that slightly younger (16-month-old) infants also failed to generalize familiar morpho-syntactic dependencies in natural languages to unfamiliar intervening elements (nonce words), unless the latter alternated with familiar intervening elements (real words). Thus, if the results in Experiment 1 Chapter 4 do stem from the inability of 18-month-olds to generalize (and not from methodological issues obscuring the existence of an effect), this would still align with natural language acquisition. Failure to generalize may be caused by a variety of factors, some of which discussed in the previous section. Importantly, however, previous research as well as the result of Experiment 2 in Chapter 4 supports the conclusion that it is not caused by rote-learning, as both adults and infants learn dependencies best in conditions of high variability in the intervening element. If infants are sensitive to the specific one-to-one mapping between *a* and *b* items, but require high variability in the *X* elements then they too must be abstracting away from the specific *X* elements (but not the specific *a/b* ones).

To conclude the current section, both adults and infants are endowed with the ability to detect dependencies between remote elements by abstracting away from the intervening material. This ability could be crucial to the detection of morpho-syntactic dependencies in natural languages. In the current study I offered preliminary evidence that the properties of morpho-syntactic dependencies (and of their dependent elements, namely functors) may facilitate the ability known as NAD-learning, and thus make it a reliable mechanism for acquisition. These findings need to be extended by future research exploring the boundaries of NAD-learning, and also the specific timeline of the acquisition of morpho-syntactic dependencies in (various) natural languages.

5.5 What We Have Learned About Learning NADs

The second aim of the studies in this dissertation was to deepen our understanding of the NAD-learning mechanism. In the experiments presented here we saw a highly constrained learning process: participants could only learn NADs under specific conditions, where the dependent elements were marked by sufficient cues, and where the participants were not distracted by a preceding learning phase (unless clear cues marked the transition). The limitations of this learning mechanism can bring us closer to understanding its nature. For instance, NAD-learning has been claimed to be a distributional learning mechanism, capitalizing on the high probability of co-occurrence between the dependent elements. However, prosodic cues to dependent elements can guide dependency-learning (Chapter 2), showing that when acquiring NADs, learners may not simply compute co-occurrence probabilities indiscriminately. Instead, they may use cues in the input (and principles such as Gestalt perception for organizing the input) to first limit the number of co-occurrence relationships over which statistics may be computed.

Another basic observation about NAD-learning as derived from the adult studies in Chapters 2 and 3 is that it can take place in conditions of incidental learning, not only without explicit instructions but also while completing a simultaneous task. Like Saffran et al. (1997), I employed a colouring task that would distract learners from explicitly attending to the input and developing and testing hypotheses about its structure. Furthermore, at familiarization I did not instruct participants as to the existence of a test phase, to avoid motivating learners to find patterns in the input. Previous NAD-learning studies did not provide learners with distracting tasks, and often informed them of a subsequent test phase where they would have to demonstrate their knowledge of the language they had heard.

Therefore, the finding that NAD-learning can proceed in incidental learning conditions is in itself important. Dienes & Berry (1997) proposed that incidental learning conditions promote implicit learning, a type of learning that is unconscious, involuntary, and robust in the face of time or lack of attentional resources. However, implicit learning is linked to implicit knowledge, a sum of representations (often obtained as a result of implicit learning) that the learner is not aware of and cannot verbally reproduce. If NAD-learning in my experiments had yielded implicit knowledge of the dependencies, none of the participants should have reported explicit awareness of the rules that were being tested. On the contrary, explicit learners were frequent (and showed the highest accuracy rates) especially in the conditions were learners' performance was significantly above chance.

Thus, even though learning was incidental, the result of this process was often in the form of explicit knowledge. Furthermore, explicit learners often reported becoming aware of the regularities *during* the familiarization phase. Reber (1989) argued that implicit learning was effective with the types of grammars that were complex enough that recalling their rules explicitly would be very difficult. I proposed that, because learners in these experiments were tasked with learning a very simple rule, the more successful learners were also the ones that inevitably gained awareness of this simple rule. However, it is as yet unclear whether the participants who gained no explicit knowledge also learned less than the ones that did. Future research could use methods more sensitive to implicit knowledge alone, such as measuring reaction times (e.g. to the *b* element when it is predicted correctly or not by the *a* element in an *aXb* string) or neurophysiological response in an EEG paradigm to compare between implicit and explicit learners, and identify whether they apply different learning strategies, or whether they simply differ in the extent of learning.

Whatever might be the underlying cause for the emergence of explicit knowledge, it is clear that attention was not entirely captured by the distraction task in these experiments. Perhaps because the distractor task was in the visual modality, whereas the learning task was in the auditory modality, learners were able to engage in the distractor task while at the same time attending to the spoken input. In fact, the role of attention has been highlighted before and has been argued to be crucial to NAD-learning (Pacton & Perruchet, 2008; de Diego-Balaguer & López-Barroso, 2010; Pacton, Sobaco & Perruchet, 2015; de Diego-Balaguer et al., 2016; López-Barroso et al., 2016). The results of the current studies also support the conclusion that attention is an important aspect of tracking dependencies between non-adjacent elements.

Firstly, as shown in Chapter 3, attentional resources at familiarization modulated learners' ability to detect the regularities presented in this learning phase. Thus, learners needed explicit cues to recover their attention at the start of familiarization if a pre-familiarization phase preceded it. Furthermore, one likely explanation for the results of Experiment 1 in Chapter 3 is that the length of the learning phase before familiarization (with the target language) also negatively affects the learning performance at familiarization. It seems that the best time to learn from the input is at the beginning of an experiment, when attentional resources are not depleted and when the input is still unfamiliar enough as to engage the learner. Therefore, even in the presence of a secondary task which discourages explicit pattern-searching, some level of attention is still necessary in order to tune into the regularities of the *aXb* language.

Secondly, my results support the claim that NAD-learning is possible when there is joint attention to the dependent elements, and when (selective) attention shifts away from the intervening material (Pacton & Perruchet, 2008). In Chapter 2 I showed that prosodic cues that marked the *a/b* elements but not the *X* elements facilitated NAD-learning. Furthermore, learning seemed slightly superior when the *a/b* words were more prosodically salient than the intervening *X*. In Chapter 4, 18-month-olds were unable to track dependencies that spanned novel *X* elements at test, consistent with previous research that showed that novelty in the intervening material distracted infants from tracking dependencies they were otherwise familiar with. It must also be pointed out that learning was possible with both infants and adults in a language where the *a/b* elements were distinguished by their high frequency and low variability (see Gómez, 2002) as well as their presence in string-peripheral positions, which are perceptually salient positions (Endress et al., 2005). These additional cues marking the *a/b* elements as salient might further have served to direct the learner's attention to the target dependent elements.

The necessity of processing dependent elements jointly and shifting attention to them (and away from the intervening material) falls neatly in line with the proposal that NAD-learning is supported by Gestalt principles of similarity. If these Gestalt principles group together dependent elements with similar properties, then this facilitates their joint processing. Note that additional cues are necessary to shift attention to the dependency and away from the intervening material, and as we have seen in Experiment 1 in Chapter 4, this shift may occur in the opposite direction as well. Learning dependencies between perceptually salient elements is easily explained, as learners naturally orient their attention to perceptually salient stimuli. But why would learners direct their attention to perceptually less salient stimuli such as functors in natural languages? De Diego-Balaguer et al. (2016) point out that aside from exogenous attention, which is captured by salience cues, there is also endogenous attention, which prompts learners to selectively attend to aspects in the input based on their prior experience.

It could be, therefore, that learners direct their attention to frequent and perceptually non-salient elements because their prior linguistic experience tells them these types of elements are functional words that define the syntactic structure of sentences.

In light of this, I propose that Gestalt principles of similarity, along with selective attention, exploit cues at every level of input processing (i.e. prosodic cues, distributional cues, category membership, etc.) to highlight dependencies between functional morphemes and help identify potential morpho-syntactic relationships. Thus, as mentioned in the previous section, Gestalt principles of similarity may group together not only elements that are perceptually similar, but also elements that share distributional properties (e.g. which are equally frequent or variable). Furthermore, prior experience may direct learners to shift attention to these dependent elements because of, for instance, their frequency in the input. Once selective attention has highlighted these elements and Gestalt principles of similarity have linked them together, a statistical learning mechanism can be applied to compute co-occurrence probabilities between them, selecting among potential dependencies to favor the ones that exhibit a high transitional probability (see Van den Bos, Christiansen & Mysiak, 2012).

Adopting this perspective would explain the fact that a variety of cues (perceptual, distributional, positional etc.) can contribute to successful NAD-learning. Furthermore, learning is all the more robust if various cues (which on their own may be insufficient) were *integrated*. For instance, if both distributional and prosodic analyses of the input concurred to highlight functional morphemes as similar to each other and distinct in the environment, this would facilitate NAD-learning more than if just one type of cue (e.g. distributional) marked dependencies in the input. This in fact is the case in the study of Chapter 2, where all participants hear a language with highly frequent/invariable *a/b* elements but only those that benefit from the additional prosodic cue show successful learning. On the other hand, perhaps dependencies where the dependent elements are not marked by prosodic similarity cues (such as the French *Vous chantez*) might benefit from other cues to compensate (e.g., because functors in French have been claimed to be salient than in other languages like English, - see Shi et al., 2006 – these morphemes [*vous, -ez*] may be encoded and categorized sooner, and prior familiarity with their properties might direct attention to them in the input).

The current proposal makes several specific predictions. Firstly, that even when dependencies are learned entirely implicitly and learners remain unaware of the existence of these patterns, learning may still be conditioned by Gestalt principles of similarity and the existence of cues that direct attention to the dependencies and away from the intervening material (Pacton & Perruchet, 2008). Therefore, if a more complex grammar involving non-adjacent patterns were tested using more indirect methods suited to gleaning implicit learning (e.g. reaction time tests or EEG), learning would be supported by the same type of cues identified in the current study. A second prediction is that prior familiarity with the dependent elements and their properties, and especially with the class-distinction between *a/b* and *X* elements should facilitate the detection of dependencies in *aXb* strings. Similarly, infants who have acquired functional morphemes (both words and affixes) as a category distinct from lexical ones should find it easier to learn dependencies between these morphemes. Thus, further research continuing the investigation in Chapter 3 should yield a positive effect of pre-familiarization, confirming the hypothesis that Starting Small is better than Starting Big, where NAD-learning is concerned.

To conclude the current section, the results of this study have furthered our understanding of the constraints that limit the working of the NAD-learning mechanism, and have shown that these constraints align with Gestalt principles of similarity. The findings have also added to a growing body of research that investigates the role of attention in NAD-learning, and have confirmed that learners of dependencies need to be guided to attend selectively to the elements in the input that can enter such dependencies. Together, these results point to a learning mechanism that requires dependent elements to be simultaneously highlighted in the input before their high co-occurrence probabilities are detected by the learner.

5.6 Methods and Questions for Future Research

In this last section I discuss a few directions that I believe should be taken in future research on NAD-learning. I focus specifically on methodological issues, and how innovative methodology can give us the kind of insight we require into the workings of this mechanism.

In previous sections we saw that different methods of testing in an AGL paradigm can tap into different types of representations that learners have obtained from exposure. While participants' rating of individual test items as in/consistent with exposure (the method used here) might reflect explicit learning, participants' online reaction upon encountering a(n) in/consistent dependent element ($a_iX\underline{b_i}$ vs. $a_iX\underline{b_j}$) might reflect more implicit knowledge. Previous work has shown that adult participants do show a difference in reaction times to correct or incorrect dependencies (Mysiak, Christiansen & Tomblin, 2010; de Vries et al., 2012; López-Barroso et al., 2016), indicating that they use the first dependent element to predict the occurrence of the last one (in real time). However very few studies have directly compared implicit and explicit knowledge derived from NAD-learning (though see López-Barroso et al., 2016), to identify whether they are sensitive to the same types of cues. It is important to determine whether obtaining implicit representations of NADs is also conditioned by the presence of sufficient cues directing attention to the dependent elements and the relationship between them.

Another important project for future research would be to identify the exact type of statistical computation that underlies NAD-learning. While most previous NAD-learning studies tested dependencies which exhibited 100% (backward and forward) transitional probability between dependent elements, Van den Bos, Christiansen & Mysiak (2012) tested learning with 'deterministic' as well as 'probabilistic' dependencies. Probabilistic dependencies were instantiated between an initial a_i element that combined half the time with a b_i and the rest of the time with a different b_j element. When dependencies were probabilistic (i.e. with 50% forward transitional probability), additional phonological cues were necessary to make them learnable. While Van den Bos et al.'s study certainly shows that learners are sensitive to co-occurrence probabilities, and show superior learning when these probabilities are higher, more research is needed to determine the exact nature of the statistical computation that learners perform. For instance, would learners be equally sensitive to the backward transitional probability (e.g if a_i could combine only with b_i, but b_i appeared in combination with either a_i or a_j)?

Not only is it important to identify the specific type of mental computations that learners are performing, but it is also important to determine whether all learners are the same. In most AGL experiments, the significant learning effect in a condition (20-30 subjects) is usually carried by less than a third of the sample (Siegelman, Bogaerts & Frost, 2015). However, experimenters generalize these learning abilities to the entire population. Siegelman, Bogaerts & Frost (2015) proposed that a more ecologically valid approach would be to capture differences between the learners. A variety of testing methods and types of test items should reveal whether different subjects may learn different aspects of the input. For instance, while some of the learners in the present study may have focused on detecting the (item-bound) *a_b* dependencies, others may have been building categories for *a, b* or *X* elements. These differences may also correlate with individual differences in processing, (Misyak & Christiansen, 2012), cognitive style (Riding, 1991) or working memory abilities.

Finally, we must acknowledge that while most AGL literature and the present study as well speak of 'learning', this term may often simply equal 'detection', the development of a temporary sensitivity to patterns in the input. If NAD-learning is to be a reliable mechanism for language acquisition, it should produce stable representations that can be encoded in long-term memory. However, few studies to this date have investigated to what extent the learning of

NADs in an AGL experiment can stand the test of time. López-Barroso et al. (2006) found no evidence that participants who had detected NADs on a previous day retained their sensitivity to the dependencies 24 hours later. Gomez et al. (2006) and Hupbach et al. (2009) argued that infants who napped immediately after exposure retained an abstract representation of the dependencies (the notion that there is a relationships beween the first and last word in an *aXb* string without the knowledge of which *a* word predicts which *b* word specifically). This abstract representation was still present on the following day after testing. On the other hand, infants who didn't nap remembered the specific dependencies on the same day of testing, but retained neither the item-specific nor the abstract representations of the dependencies on the following day.

Sleep plays an important role in consolidating memories, and in infants up to 18 months this role seems to be linked to generalization and abstraction. In both adults and infants a 24-hour delay in testing seems to trigger a decay in memories for specific items that can enter dependencies. However, future research may reveal that with repeated familiarization over a number of subsequent days, even item-specific representations may become robust enough to be consolidated and not erased over a night's sleep.

To conclude, while many questions remain open about the specific ways in which NADs might be computed and represented in the mind of each individual learner, and how these initial representations could hypothetically grow to become part of one's solid linguistic system, current AGL research possesses a variety of interesting experimental methods that can guide us to better understanding the process of NAD-learning.

Bibliography

Allen, R., & Reber, A. S. (1980). Very long term memory for tacit knowledge. *Cognition*, 8(2), 175-185.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, 77(6), 1698-1716.

Arnon, I. (2009). *Starting Big – The Role of Multi-Word Phrases In Language Learning And Use*. Doctoral Dissertation, Stanford University.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, 122(3), 292-305.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning from acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21(3), 170-176.

Aslin, R.N., Saffran, J.R. & Newport, E.L. (1998). Computation of conditional probability statistics by human infants. *Psychological Science*, 9, 321-324.

Austin, A., & Newport, E. L. (2011). When children learn more than they are taught: Regularization in child and adult learners. Unpublished manuscript, University of Rochester, Rochester, NY.

Bassano, D., Eme, P. E., & Champaud, C. (2005). A naturalistic study of early lexical development: General processes and inter-individual variations in French children. *First Language*, 25(1), 67-101.

Berendson, E. (1986). *The Phonology of Cliticization*. Ph.D. dissertation, Utrecht.

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.3.03) [Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/

Booij, G. (1996). Cliticization as prosodic integration: the case of Dutch. *The Linguistic Review*, 13(3-4), 219-242.

van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, 67(4), 507-520.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, 27(6), 807-842.

Braine, M. D. S., Brody, R. E., Brooks, P. D., Sudhalter, V., Ross, J. E., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language*, 29, 591–610.

Brown, R. & Fraser, C. (1963). The Acquisition of Syntax. In Bellugi, U. & Brown, R. (Eds.) *The Acquisition of Language* (pp. 43-79). University of Chicago Press.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, 121(1), 127-132.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33(3), 386-404.

Chomsky, N. (1980). *Rules and representations*. Columbia University Press.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. MIT press.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, 2(10), 406-416.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120(3), 235.

Cochran, B. P., McDonald, J. L., & Parault, S. J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language*, 41(1), 30-58.

Conway, C. M., Ellefson, M. R., & Christiansen, M. H. (2003). When less is less and when less is more: Starting small with staged input. In Alterman, R. & Kirsh D. (Eds.) *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 270-275). Mahwah, NJ: Lawrence Erlbaum.

Creel, S.C., Newport, E.L. & Aslin, R.N. (2004). Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30 (5), 1119-1130.

Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of

spoken sentences. *Journal of Psycholinguistic Research*, 22, 109-131.

Dempster, F. N. (1978). Memory span and short-term memory capacity: A developmental study. *Journal of Experimental Child Psychology*, 26, 419–431.

de Diego‑Balaguer, R., Fuentemilla, L. & Rodriguez-Fornells, A. (2011). Brain dynamics sustaining rapid rule extraction from speech. *Journal of Cognitive Neuroscience*, 23(10), 3105-3120.

de Diego‑Balaguer, R., & Lopez‑Barroso, D. (2010). Cognitive and neural mechanisms sustaining rule learning from speech. *Language Learning*, 60(s2), 151-187.

de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology*, 7:44.

de Diego-Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Lévi, A. C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS One*, 2(11), e1175.

Dienes, Z. (2007). Subjective measures of unconscious knowledge. *Progress in brain research*, 168, 49-64.

Dienes, Z., & Berry, D. (1997). Implicit learning: below the subjective threshold. *Psychonomic Bulletin & Review*, 4(1), 3-23.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71-99.

Endress, A. D. (2010). Learning melodies from non-adjacent tones. *Acta Psychologica*, 135(2), 182-190.

Endress, A.D. & Bonatti, L.L. (2006). Rapid learning of syllable classes from a perceptually continuous stream. *Cognition*, 105 (2), 247-299.

Endress, A.D.& Mehler, J. (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology*, 62 (11), 2187-2209.

Endress, A.D. & J. Mehler (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 235-250.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*, 134(3), 406.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 458.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99(24), 15822-15826.

Friederici, A.D., Mueller, J.L. & Oberecker, R. (2011). Precursors to natural grammar learning: preliminary evidence from 4-month-old infants. *PLoS One*, 6(6): e20726.

Gathercole, S. E., Pickering, S.J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2), 177.

Geambasu, A. & Levelt, C.C. (2015, June). Failures and successes in infant artificial grammar learning. Poster presented at the *Interdisciplinary Advances in Statistical Learning* Conference, Basque Centre on Cognition, Brain and Language, San Sebastian, Spain. Abstract retrieved from http://www.bcbl.eu/events/statistical-learning/en/conference/

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*, 33(6), 1087-1116.

Gebhart, A.L., Newport, E.L. & Aslin, R.N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, 16 (3), 486-490.

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, 98(3), B67-B74.

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, 26(2), 204.

Gerken, L. A.,Wilson, R., & Lewis,W. (2005). Infants can use distributional cues to form

syntactic categories. *Journal of Child Language*, 32, 249-268.

Goldberg, A.E. (2006). *Construction at Work: The Nature of Generalization in Language*. Oxford University Press.

Goldowsky, B., & Newport, E. L. (1990). The less is more hypothesis: modeling the effect of processing constraints on language learnability. Unpublished manuscript, University of Rochester, Rochester, NY.

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

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science*, 17(8), 670-674.

Gómez, R. L., & Edgin, J. O. (2015). Sleep as a window into early neural development: Shifts in sleep‐dependent learning effects across early childhood. *Child Development Perspectives*, 9(3), 183-189.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70(2), 109-135.

Gómez, R.L. & Lakusta, L. (2004). A first step in form-based category abstraction by

12-month-old infants. *Developmental Science*, 7 (5), 567-580.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7, 183-206.

Gómez, R. L., & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(2), 396.

Grama, I.C., Wijnen, F.N.K & Kerkhoff, A.O. (2013). Constraints on non-adjacent dependency-learning: distance matters. In Baiz, S., Goldman, N. & Hawkes, R., (Eds), *BUCLD 37 Online Proceedings Supplement*. http://www.bu.edu/bucld/supplementvol37/

Green, T. R. (1979). The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior*, 18(4), 481-496.

Hall, T. A. (1999). Phonotactics and the prosodic structure of German function words. *Amsterdam Studies in the Theory and History of Linguistic Science Series*, 4, 99-132.

Hallé, P.A., Durand, C. & de Boysson-Bardies, B. (2008). Do 11-month-old French infants process articles? *Language and Speech*, 51 (1&2), 23-44.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598), 1569-1579.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, 78(3), B53-B64.

Hazeltine, E., Grafton, S. T., & Ivry, R. (1997). Attention and stimulus characteristics determine the locus of motor-sequence encoding. A PET study. *Brain*, 120(1), 123-140.

van Heugten, M. & Johnson, E.K. (2010). Linking infants'distributional learning abilities to natural language acquisition. *Journal of Memory and Language*, 63, 197-209.

van Heugten, M. & Shi, R. (2010). Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *The Journal of the Acoustical Society of America*, 128 (5), EL223 – EL228.

Höhle, B., & Weissenborn, J. (2000). The origins of syntactic knowledge: Recognition of determiners in one year old German children. In Howell, C., Fish, S.A. & Keith-Lucas, T. (Eds.) *Proceedings of the 24th Annual Boston University Conference on Language Development, Volume 2* (pp. 418-429). Sommerville, MA: Cascadilla Press.

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, 6(2), 122-127.

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, 5(3), 341-353.

Höhle, B., Schmitz, M., Santelmann, L. M. & Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities. *Language Learning and Development*, 2 (4), 277-300.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, 13(4), 341-348.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2), 151-195.

Hudson Kam, C. L, & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30-66.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In Rovee-Collier, C. & Lipsitt, L. P. (Eds.), *Advances in Infancy*

Research (vol. 5, pp. 69–93), Norwood, NJ: Ablex.

Hupbach, A., Gomez, R. L., Bootzin, R. R., & Nadel, L. (2009). Nap-dependent learning in infants. *Developmental Science*, 12(6), 1007-1012.

Johnson, E. K. (2005). English-learning infants' representations of word forms with iambic stress. *Infancy*, 7(1), 99-109.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44(4), 548-567.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, 13(2), 339-345.

Jun, S.-A. & Fougeron, C. (2000). A Phonological model of French intonation. In Botinis, A. (ed.), *Intonation: Analysis, Modeling and Technology*, (pp. 209-242). Dordrecht: Kluwer Academic Publishers.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General*, 126(3), 278.

Kedar, Y., Casasola, M., & Lust, B. (2006). Getting there faster: 18-and 24-month-old infants' use of function words to determine reference. *Child Development*, 77(2), 325-338.

Kemler-Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, 18(1), 111-116.

Kerkhoff, A.O., de Bree, E., de Klerk, M. & Wijnen, F. N.K. (2013). Non-adjacent dependency learning in infants at familial risk of dyslexia. *Journal of Child Language*, 40 (1), 11-28.

Kersten, A. W., & Earles, J. L. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language*, 44(2), 250-273.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35-B42.

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language*, 66(1), 285-306.

Lany, J. & Gómez, R.L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, 19 (12), 1247-1252.

Lidz, J., Omaki, A., & Orita, N. (2012). Input in the lab causes 15-month-olds to learn an English non-adjacent dependency. Paper presented at *BUCLD 37*, Boston University. Abstract retrieved from: http://www.bu.edu/bucld/files/2013/07/2012-handbook-web-version.pdf

López-Barroso, D., Cucurell, D., Rodríguez-Fornells, A., & de Diego-Balaguer, R. (2016). Attentional effects on rule extraction and consolidation from speech. *Cognition*, 152, 61-69.

Ludden, D., & Gupta, P. (2000). Zen in the art of language acquisition: Statistical learning and the less is more hypothesis. In Gleitman, L.R. & Joshi, A. K. (Eds.) *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology*, 67(3), 130-150.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398), 77-80.

Marquis, A., & Shi, R. (2012). Initial morphological learning in preverbal infants. *Cognition*, 122(1), 61-66.

Maye, J., & Gerken, L. (2000, March). Learning phonemes without minimal pairs. In Howell, S.C. et al. (eds), *Proceedings of the 24th Annual Boston University Conference on Language Development, Volume 2* (pp. 522-533). Sommerville, MA: Cascadilla Press.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101-B111.

McCarthy, J. L. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, 12, 373-418.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, 30(5), 678-686.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91-117.

Mintz, T. H. (2006). Finding the verbs: Distributional cues to categories available to young learners. In Hirsh-Pasek, K. & Golinkoff, R. M. (eds.), *Action meets word: How*

children learn verbs (pp. 31-63). Oxford: Oxford University Press.

Mintz, T. H. (2013). The segmentation of sub-lexical morphemes in english-learning 15-month-olds. *Frontiers in Psychology*, 4(24).

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62, 302–331.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, 2, 138–153.

Mitchell, A. & Weiss, D.J. (2010). What's in a face? Visual contributions to speech segmentation. *Language and Cognitive Processes*, 25(4), 456-482.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential contribution of phonological and distributional cues in grammatical categorization. *Cognition*, 96, 143–182.

Monaghan, P., Christiansen, M. H. & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, 55, 259-305.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19(4), 498-550.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In Morgan, J.L. & Demuth, K. (eds), *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, (pp. 263-283). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2010). Learnability of embedded syntactic structures depends on prosodic cues. *Cognitive Science*, 34(2), 338-349.

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, 109(39), 15953-15958.

Nazzi, T., Barrière, I., Goyet, L., Kresh, S., & Legendre, G. (2011). Tracking irregular morphophonological dependencies in natural language: Evidence from the acquisition of subject-verb agreement in French. *Cognition*, 120(1), 119-135.

Nelson, C. A. (1995). The ontogeny of human memory: A cognitive neuroscience perspective. *Developmental Psychology*, 31(5), 723.

Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences*, 10(1), 147-172.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14(1), 11-28.

Newport, E. L. & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127-162.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, 49(2), 85-117.

Onnis, L., Monaghan, P. , Christiansen, M. H. & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Erlbaum: Mahwah, NJ, 1047-1052.

Onnis, L., Monaghan, P., Richmond, K. & Chater, N. (2005) Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, 53, 225-237.

Pelphrey, K. A., Reznick, J. S., Davis Goldman, B., Sasson, N., Morrow, J., Donahoe, A., & Hodgson, K. (2004). Development of visuospatial short-term memory in the second half of the 1st year. *Developmental Psychology*, 40(5), 836.

Peña, M., Bonatti, L. L., Nespor, M. & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604-607.

Pacton, S., Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34(1), 80-96.

Pacton, S., Sobaco, A., & Perruchet, P. (2015). Is an attention-based associative account of adjacent and nonadjacent dependency learning valid? *Acta Psychologica*, 157, 195-199.

Pavlidou, E. V., Kelly, M.L., & Williams, J. M. (2010). Do children with developmental dyslexia have impairments in implicit learning? *Dyslexia*, 16(2), 143-161.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: no need for algebraic-like computations. *Journal of Experimental Psychology: General*, 133(4), 573.

Perruchet, P., Vinter, A., Pacteau, C., & Gallego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *The Quarterly Journal of Experimental Psychology*, 55A(2), 485-503.

Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.

Peterson, E.R., Deary, J.I. & Austin, E.J. (2003). The reliability of Riding's Cognitive Styles Analysis test. *Personality and Individual Differences*, 34, 899-904.

Pinker, S. (1998). Words and rules. *Lingua*, 106(1), 219-242.

Ravignani, A., Sonnweber, R.S., Stobbe, N., & Fitch, W.T. (2013). Action at a distance: dependency sensitivity in a New World primate. *Biology Letters*, 9(6), 20130852.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81(1), 115.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General,* 118 (3), 219-235.

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition*, 6(3), 189-221.

Reber, P. J., & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience*, 10(2), 248-263.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: the role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, 66, 30-54.

Riding, R. (1991). *Cognitive Styles Analysis*. Birmingham: Learning and Training Technology.

Riding, R. & Rayner, S.G. (1998). *Cognitive Styles and Learning Strategies: Understanding Style Differences in Learning and Behavior*. London: David Fulton.

Riding, R., & Sadler-Smith, E. (1992). Type of instructional material, cognitive style and learning performance. *Educational Studies*, 18(3), 323-340.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Visual short-term memory in the first year of life: Capacity and recency effects. *Developmental Psychology*, 37(4), 539.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44(4), 493-515.

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27-52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606-621.

Saffran, J.R., Newport E.L., Aslin, R.N., Tunick, R.A. & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science*, 8(2), 101-105.

Santelmann, L. M. & P.W. Jusczyk (1998). Sensitivity to discontinuous dependencies in language learners: evidence for the limitations in processing space. *Cognition*, 69, 105-134.

Selkirk, E.O. (1986). On derived domains in sentence phonology. *Phonology Yearbook,* 3, 371-405.

Selkirk, E.O. (1996). The prosodic structure of function words. In Morgan, J.L. & Demuth, K. (eds), *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, (pp. 187–213). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Santelmann, L., Jusczyk, P., & Huber, M. (2003). Infants Attention to Affixes. *Jusczyk Lab Final Report, edited by D. Houston, A. Seidl, G. Hollich, E. Johnson, and A. Jusczyk.* http://hincapie.psych.purdue.edu/Jusczyk

Shady, M. E. (1996). *Infants' Sensitivity to Function Morphemes*. Unpublished doctoral dissertation. The State University of New York, University at Buffalo.

Shafer, V. L., Shucard, D. W., Shucard, J. L., & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language, and Hearing Research*, 41(4), 874-886.

Shi, R., Cutler, A., Werker, J. & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of the Acoustical Society of America*, 119 (6), EL61-EL67.

Shi, R. & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*, 11 (3), 407-413.

Shi, R., Marquis, A., Gauthier, B. (2006). Segmentation and representation of function words in preverbal French-learning infants. In Bamman, D., Magnitskaia, T., & Zaller, C. (Eds.) *Proceedings of the 30th Annual Boston University Conference on Language Development Volume 2* (pp. 549-560). Somerville, MA: Cascadilla Press.

Shi, R., & Melançon, A. (2010). Syntactic Categorization in French‐Learning Infants. *Infancy*, 15(5), 517-533.

Shi, R., Morgan, J.L. & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, 25, 169-201.

Shi, R. & Werker, J. (2001). Six-month-old infants' preference for lexical words. *Psychological Science*, 12 (1), 70-75.

Shi, R. & Werker, J. (2003). The basis of preference for lexical words in 6-month-old infants. *Developmental Science*, 6 (5), 484-488.

Shi, R., Werker, J. & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, 10 (2), 187-198

Shi, R., Werker, J. & Morgan, J.L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72, B11-B21

Siegelman, N., Bogaerts, L. & Frost, R. (2015, June). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. Paper presented at the *Interdisciplinary Advances in Statistical Learning* Conference, Basque Centre on Cognition, Brain and Language, San Sebastian, Spain. Abstract retrieved from http://www.bcbl.eu/events/statistical-learning/en/conference/

Soderstrom, M. (2002). *The Acquisition of Inflection Morphology In Early Perceptual Knowledge Of Syntax*. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore.

Soderstrom, M., White, K. S., Conwell, E., & Morgan, J. L. (2007). Receptive grammatical knowledge of familiar content words and inflection in 16-month-olds. *Infancy*, 12(1), 1-29.

Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition*, 18(3), 733-745.

Strömqvist, S., Ragnasdóttir, H. & Richthoff, U. (2001). Input and production in the earliest development of function words. In Weissenborn, J. & Höhle, B. (Eds), *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition* (vol. 2, pp. 157-178). Amsterdam: Benjamins.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology*, 39(4), 706.

Tincoff, R., Santelmann, L., & Jusczyk, P. (2000). Auxiliary verb learning and 18-month-olds' acquisition of morphological relationships. In Howell, S.C., Fish, S.A. & Keith-Lucas, T. (eds.), *Proceedings of the 24th annual Boston University*

Conference on Language Development (Vol. 2, pp. 726-737). Somerville, MA: Cascadilla Press.

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

Udden, J., Ingvar, M., Hagoort, P., & Petersson, K. M. (2012). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: Investigating the Push-Down Stack Model. *Cognitive Science*, 36(6), 1078-1101.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4(02), 105-122.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/ procedural model. *Cognition*, 92(1), 231-270.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, 27(1), 71-86.

Vigário, M. (1999). On the prosodic status of stressless function words in European Portuguese. *Amsterdam Studies in the Theory and History of Linguistic Science Series*, 4, 255-294.

de Vries, M. H., Petersson, K. M., Geukes, S., Zwitserlood, P., & Christiansen, M. H. (2012). Processing multiple non-adjacent dependencies: evidence from sequence learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598), 2065-2076.

Vuong, L., Meyer, A. S., & Christiansen, M. H. (2011). Simultaneous online tracking of adjacent and non-adjacent dependencies in statistical learning. In Carlson, L. , Hölshcer, C. & Shipley, T. (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 964–969). Austin, TX: Cognitive Science Society.

Weiss, D.J., Gerfen, C. & Mitchell, A. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development*, 5, 30-49.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, 53(1), 67-121.

Wilsenach, C. & Wijnen, F.N.K. (2004). Perceptual sensitivity to morphosyntactic agreement in language learners: evidence from Dutch children at risk for developing dyslexia. In A. Brugos, L. Micciulla & C.E. Smith (eds), *BUCLD 28: Proceeding of the 28th Annual Boston University Conference on*

Language and Development, Volume 2 (pp. 645-656). Sommerville, MA: Cascadilla Press.

Zec, D. (2002). On the prosodic status of function words. *Working Papers of the Cornell Phonetics Laboratory*, 14, 207-249.

Zellner, B. (1994). Pauses and the temporal structure of speech. In Keller, E. (Ed.) *Fundamentals of Speech Synthesis and Speech Recognition* (pp. 41-62). Chichester: John Wiley.

Zink, I. & Lejaegere, M. (2002). *N-CDIs : Lijsten voor communicatieve ontwikkeling* [NCDIs: Lists for communicative development]. Leuven: Acco.

Zinszer, B. D., & Weiss, D. J. (2013). When to hold and when to fold: detecting structural changes in statistical learning. In Knauff, M., Paunen, M., Sebanz, N., Wachsmuth, I. (Eds.), *Proceedings of the Thirty-Fifth Annual Conference of the Cognitive Science Society* (pp. 3858-3863). Austin, TX: Cognitive Science Society.

Appendix

Tables for Chapter 2

		EMPHASIZED			LEXICAL			FUNCTIONAL		
		Mean pitch (Hz)	Mean amplitude	Duration (s)	Mean pitch (Hz)	Mean amplitude	Duration (s)	Mean pitch (Hz)	Mean amplitude	Duration (s)
	TEP **[tɛp]**	314.5	81.43	0.28	233.6	78.69	0.3	199	74.22	0.16
	SOT **[sɔt]**	289.6	77.59	0.5	232.7	79.75	0.43	198.1	75.71	0.21
a	**RAK** **[rɑk]**	276.3	77.55	0.41	234.7	77.06	0.38	198.8	76.48	0.18
	LUT **[lœt]**	289.1	81.82	0.47	234.1	79.2	0.38	200.2	78.27	0.17
b	**JIK** **[jIk]**	293.8	81.27	0.43	232.7	79.46	0.37	197.8	78.89	0.18
	TOEF **[tuf]**	279.7	80.53	0.48	239.6	77.45	0.38	198.1	77.27	0.16
	Mean	**290.5**	**80.03**	**0.428**	**234.5**	**78.6**	**0.373**	**198.7**	**76.81**	**0.177**

Table A2.1. The *a/b* tokens, with IPA transcriptions and acoustic measures in Experiment 1 [14]

[14] To avoid large variations between acoustic properties of a/b tokens in a Condition, I resynthesized some of these tokens to match them for pitch and duration (the standard values were taken as the mean values of the original recordings): one element was

No.	X		Mean pitch (Hz)	Mean pitch of the first (stressed) syllable (Hz)	Mean intensity (dB)	Duration (s)	Duration of the first syllable (s)
Familiarization		*IPA*					
1	blieker	**[blikər]**	192.7	247.1	79.27	0.54	0.23
2	dufo	**[dyfo]**	166.1	250.2	79.99	0.57	0.2
3	fidang	**[fidɑŋ]**	199.5	260.6	78.73	0.62	0.11
4	gopem	**[xopəm]**	173.4	247.7	79.29	0.65	0.18
5	kengel	**[kɛŋəl]**	198.1	252.3	80.06	0.41	0.25
6	kijbog	**[kɛibɔx]**	196.4	203.6	80.6	0.55	0.2
7	loga	**[loxa]**	182.1	244.5	79.89	0.55	0.27
8	malon	**[malɔn]**	195.9	255.9	81.78	0.61	0.24
9	movig	**[movɪx]**	198.3	233.5	80.17	0.65	0.28
10	naspu	**[nɑspu]**	180.3	272.7	79.41	0.55	0.27
11	nijfoe	**[nɛifu]**	197.8	261.7	81.41	0.6	0.28
12	noeba	**[nuba]**	197.7	260	80.1	0.53	0.25
13	plizet	**[plizɛt]**	191	253.9	80.71	0.59	0.26
14	rajee	**[raje]**	201.3	255	78.99	0.59	0.19
15	rogges	**[rɔxəs]**	203.9	265.8	78.44	0.6	0.21
16	seeta	**[seta]**	179.4	254.4	77.56	0.6	0.27
17	snigger	**[snɪxər]**	175.1	265.3	76.56	0.58	0.26
18	wabo	**[ʋɑbo]**	198.8	240	81.53	0.59	0.21
	mean		**190.43**	**251.34**	**79.69**	**0.57**	**0.23**
	Test						
19	nilbo	**[nIlbo]**	204.9	236.9	81.46	0.65	0.19
20	pergon	**[perxɔn]**	172.7	275	77.79	0.62	0.34

Table A2.2. The intervening *X* elements, used in Experiments 1 & 3

shortened (i.e. functional *jik*, from 0.21s to 0.18s) and 10 out of 12 were modified in pitch (four of the lexical variants and all of the functional variants). Elements in the Emphasized and Lexical Conditions were also scaled to an absolute peak of 0.99, whereas Functional *a/b*s were scaled to 0.85, resulting in lower amplitude values for Functional a/bs.

		NEW FUNCTIONAL Condition		
		Mean pitch (Hz)	Mean amplitude (dB)	Duration (s)
a	**TEP**	219.9	80.4	0.176
	SOT	220.8	77.3	0.277
	RAK	209.8	76	0.25
b	**LUT**	188.9	78.7	0.305
	JIK	184.4	79.2	0.27
	TOEF	191.6	74.3	0.346
	Mean	202.6	77.7	0.2707

Table A2.3. Acoustic measures of the *a/b* tokens for Experiment 2.[15]

[15] Note that while other measurements are comparable to the old recordings, the duration of the new tokens is longer, particularly of the *b* tokens - this lengthening is a natural consequence of their phrase-final position.

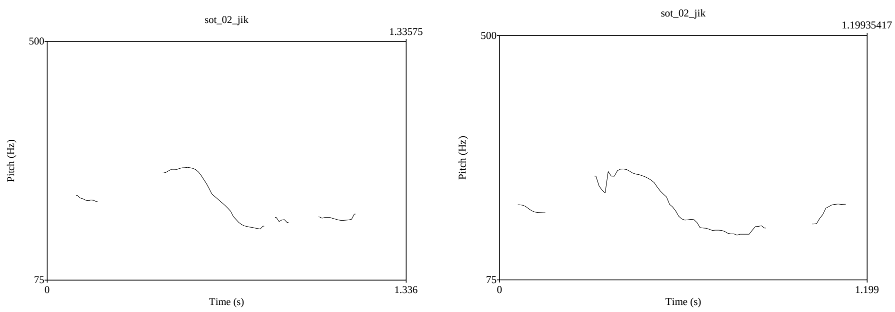

Figure A2.1 Example of the pitch contour of the same *aXb* string in the New Functional Condition (100 ms) in Experiment 2 (left) and the Old Functional Condition (100 ms) in Eperiment 1 (right).

Tables for Chapter 4

Experiment 1					Kerkhoff et al. (2013)			
Items	**Duration (ms)**	**Intensity (dB)**	**Pitch (Hz)**	**Pitch 1st syll.**	**Duration (ms)**	**Intensity (dB)**	**Pitch (Hz)**	**Pitch 1st syll.**
a								
Tep	320	59.37	384.9	-	357	59.93	284.7	-
Sot	498	57.44	375.9	-	484	61.66	300.6	-
b								
Lut	443	59.32	378.8	-	523	60.41	306.3	-
Jik	429	57.81	362.1	-	423	60.01	292.3	-
X								
wadim	572	61.15	260	300.7	596	62.4	204.7	178.9
kasi	510	61.77	243.5	279.3	486	60.53	201.5	182.9
poemer	415	63.49	267.7	361.7	497	63.08	204.6	204.9
kengel	413	65.8	297.7	355.9	511	62.46	215.6	208.6
domo	516	64.41	257.9	315.4	584	64.57	211.4	190
loga	573	62.29	255.7	314.8	604	63.27	229.2	204.8
gopem	593	62.2	249.1	285.8	697	62.58	226.8	200.8
naspu	591	62.51	276.1	323.1	625	62.62	228.4	193.5
hiftam	603	61.26	246.4	317.5	645	62.04	229.9	241.6
dieta	402	59.05	251.2	362.6	595	59.37	198.6	191.6
vami	578	63.15	245.5	303.5				
snigger	572	58.53	233.9	353.5				

rogges	541	59.34	288.7	335.8				
densim	601	58.1	270	393.9				
fidang	607	59.36	253.5	392.4				
rajee	598	60.53	244.2	355.2				
seeta	626	59.22	211.5	262				
noeba	524	63.28	248.6	292.3				
plizet	591	61.18	245.6	356.1				
banip	594	60.95	263.5	306.2				
movig	641	62.48	240.5	293.2				
sulep	563	61.27	254	333.5				
nilbo	652	60.92	231.3	274.7				
wiffel	456	59.68	226.7	361.4				
X'								
klepin	576	59.39	259.1	346.9	-	-	-	-
lotup	600	60.02	293.6	336.3	-	-	-	-
tarsin	617	61.38	244.2	299.5	-	-	-	-
Mean (X)	**561**	**59.27**	**254**	**326**	**584**	**62.29**	**215**	**200**

Table A4.1. Acoustic measures of the *a,b,* and *X* tokens in Experiment 1

Item		Duration (ms)	Pitch (Hz)	Intensity (db)
a	**tep**	302	247.1	53.52
	sot	445	254.4	56.41
b	**lut**	346	183.4	46.41
	jik	345	185.2	44.67

Table A4.2. Acoustic measures for the functor-like *a/b* tokens in Experiment 2

Samenvatting in het Nederlands

Dit proefschrift onderzoekt het menselijke vermogen om patronen te herkennen van het samen voorkomen van niet-aangrenzende elementen in onbekende input. In eerder onderzoek is voorgesteld dat het leermechanisme dat zowel volwassenen als jonge kinderen in staat stelt om het verband tussen *a* en *b* op te merken in een *aXb*-reeks (het leren van een niet-aangrenzende afhankelijkheid = non-adjacent dependency learning, NADL) taalverwerving ondersteunt het ontdekken van een morfosyntactische afhankelijkheid als in *Ik **heb** vandaag de dokter **ge**beld*. Dit proefschrift is een verslag van mijn onderzoek naar de vraag of non-adjacent dependency learning (NADL) in principe relevant kan zijn voor de verwerving van morfosyntactische afhankelijkheden in natuurlijke talen. Hiervoor heb ik leerexperimenten gebruikt met een kunstmatige grammatica die zowel gericht zijn op volwassenen als op jonge kinderen. Ik heb getest of de specifieke eigenschappen van morfosyntactische afhankelijkheden en de temporele volgorde van hun verwerving een NADL-mechanisme zouden kunnen ondersteunen, en omgekeerd, of de eigenschappen van het NADL-mechanisme het tot een geschikt hulpmiddel voor taalverwerving zouden kunnen maken.

In mijn onderzoek heb ik drie hoofdlijnen gevolgd. Ten eerste heb ik waargenomen dat morfosyntactische afhankelijkheden tot stand worden gebracht tussen functionele morfemen en dat zij gewoonlijk voorkomen met lexicale morfemen ertussenin. Eerder werk heeft aangetoond dat functionele morfemen in veel talen worden gekenmerkt door een stelsel van perceptuele cues die hen minder opmerkelijk maaken dan hun lexicale tegenhangers. Daarentegen hebben eerdere NADL-experimenten laten zien dat niet-aangrenzende afhankelijkheden gemakkelijker te op te sporen zijn wanneer perceptuele cues de afhankelijkheidselementen als in overeenstemming met elkaar marken maar ook als onderscheiden van de element in hun directe omgeving. Ik heb daarom de voorspelling getoetst dat de perceptuele cues die functionele morfemen als afzonderlijk van lexicale morfemen markeren, de herkenning van NADs tussen functionele morfemen kunnen faciliteren.

Volwassen deelnemers die aan een taal werden blootgesteld met de structuur *aXb* zijn getest op het oppikken van van *a_b*-afhankelijkheden onder verschillende omstandigheden, waar de *a/b*-tokens perceptueel (i) saillanter (Emphasized Condition), (ii) even saillant (Lexical Condition), of (iii) minder saillant (Functional Condition) zijn dan het tussenliggende *X*-woord. Het leren wordt gekwantificeerd als een voorkeur voor correcte $a_i\_b_i$ boven incorrecte $a_i\_b_j$ afhankelijkheden. Volwassen leerders laten een significante voorkeur zien voor correcte afhankelijkheden in de Emphasized en Functional Conditions maar niet in de Lexical Condition. Bovendien is dit significante leereffect alleen te zien onder omstandigheden met een natuurlijke prosodie: onder de Emphasized en Functional Conditions alleen als de opvallende *a/b*-tokens van *X* gescheiden worden door lange pauzes van 250ms, terwijk het in de Functional Condition alleen te zien is wanneer perceptueel gereduceerde *a/b*-tokens worden gescheiden door korte, nauwelijks opmerkbare pauzes van 100ms en deze geïntegreerd zijn in een natuurlijke prosodische contour.

Deze resultaten bevestigen eerdere hypotheses die stellen dat leerders van NADs beïnvloed worden door principes van Gestalt-perceptie: zij ‘groeperen’ perceptueel vergelijkbare elementen en sporen afhankelijkheden daartussen op. Daarnaast suggereren de resultaten dat deze principes van Gestalt-perceptie wellicht ook van toepassing zijn op het niveau van prosodische analyse: afhankelijkheidsleren wordt alleen vergemakkelijkt wanneer de prosodische status van de *a/b*-elementen consequent gemarkeerd is, zowel door pauzecues als door de akoestische eigenschappen van *a/b*. Tot slot suggereren deze resultaten dat leerders van natuurlijke talen het NADL-mechanisme mogelijk gebruiken om afhankelijkheden tussen prosodisch gereduceerde eenheden (functionele morfemen) op te sporen over prosodisch opvallendere eenheden (lexicale morfemen) heen. Hierbij zou NADL dus een betrouwbaar mechanisme kunnen zijn voor het verwerven van morfosyntactische afhankelijkheden in natuurlijke talen.

Als volwassen leerders beïnvloed worden door prosodische cues, wil dit echter nog niet zeggen dat jonge leerders (die hypothetisch gezien NADL gebruiken voor taalverwerving) profijt zullen hebben van hetzelfde type cues. Om dit te testen, zijn kinderen van 18 maanden oud blootgesteld aan *aXb*-talen waar de *a/b* elementen de prosodische status van functionele morfemen hebben (prosodische clitics). Na enige gewenning (“familiarisatie”) met de taal, en een korte trainingsfase, hoorden de kinderen *aXb*-testreeksen die dan wel of niet overeenkwamen met de taal die ze eerder hadden gehoord, en hun aandacht voor beide typen stimuli werd gemeten door middel van een *Headturn Preference Procedure*. De resultaten laten over het algemeen geen voorkeur zien voor een type testreeksen boven een andere. Kinderen gebruiken de eerste test trial echter om hun voorkeur te sturen gedurende de daaropvolgende test trials: als de eerste test trial overeenkomt met de taal van familiarisatie, laten de daarop volgende trials een significante voorkeur zien voor consistente boven inconsistente reeksen. Dit type afhankelijkheid van de eerste test trial om voorkeur te sturen gedurende daaropvolgende trials is al eerder geobserveerd (Gómez, Bootzin & Nadel, 2006; Hupbach, Gómez, Bootzin & Nadel, 2009): kinderen kunnen een abstracte afhankelijkheidsrepresentatie vasthouden na familiarisatie, maar slagen er niet in om de specifieke één-op-één afstemming te onthouden tussen afhankelijke elementen. De eerste trial van de testfase dient er dus toe om kinderen opnieuw in contact te brengen op met de specifieke afhankelijkheden die opgespoord moeten worden. Kortom, zowel volwassenen als kinderen lijken in staat om NADs tussen prosodisch ongemarkeerde elementen te ontwaren over prosodisch opvallendere heen. Dit vermogen zou de herkenning van morfosyntactische afhankelijkheden in natuurlijke talen kunnen ondersteunen.

Een tweede richting in het proefschrift steunt op de observatie dat de vroege gevoeligheid van baby's voor functionele morfemen in hun moedertaal voorafgaat aan hun gevoeligheid voor morfosyntactische afhankelijkheden tussen deze morfemen. Met andere woorden, baby's lijken de individuele eenheden te verwerven voordat ze gevoelig worden voor de afhankelijkheid die deze eenheden met elkaar verbindt. Daarom heb ik in een serie van experimenten onderzocht of het NADL-mechanisme vergemakkelijkt wordt door eerdere bekendheid met individuele *a/b*-tokens, voorafgaand aan blootstelling aan volledige *aXb*-reeksen. De procedure is daarbij onveranderd gebleven: volwassen deelnemers werden blootgesteld aan een *aXb*-taal en vervolgens getest op hun gevoeligheid voor specifieke *a_b*-afhankelijkheden. Een aanvullende leerfase ("prefamiliarisatie") is echter toegevoegd voorafgaand aan bekendwording met de volledige *aXb*-taal: deelnemers werden "geprefamiliariseerd" met de *a/b*-tokens in isolatie of ingebed in korte frases van twee woorden. De eerste resultaten laten zien dat het leren afneemt als de duur of de opmerkelijkheid van de prefamiliarisatie toeneemt. Het kunnen terughalen uit het geheugen van *a/b*-woorden blokkeert later de ontdekking van NADs tussen deze tokens.

Later uitgevoerd onderzoek laat echter zien dat er een vertekening zou kunnen zijn: zelfs leerders die gedurende één minuut geprefamiliariseerd zijn met woorden die compleet ongerelateerd zijn aan de *aXb*-taal, slagen er niet in om de niet-aangrenzende *a_b*-afhankelijkheden op te sporen. Eerder onderzoek suggereert dat de opeenvolgende presentatie van twee leerfases ("prefamiliarisatie" en familiarisatie) een *primacy effect* met zich meebrengt: deelnemers letten op de input gedurende de eerste leerfase maar raken hun aandacht kwijt tijdens de tweede leerfase, tenzij er een krachtige cue is die het onderscheid markeert tussen de twee leerfases (Gebhart, Aslin & Newport, 2009; Mitchell & Weiss, 2010; Zinszer & Weiss, 2013). In het laatste experiment heb ik me gericht op precies deze kwestie: deelnemers die expliciet zijn geïnstrueerd over het bestaan van twee onderscheiden leerfases, laten een superieur leereffect zien tijdens familiarisatie ten opzichte van deelnemers die deze instructie niet hebben ontvangen. De opeenvolging van twee verschillende leerfases introduceert dus inderdaad een vertekend beeld. Desalniettemin is er geen significant verschil tussen ge-prefamiliariseerd zijn met de *a/b* doelwoorden van de *aXb* taal en ge-prefamilialiseerd zijn met woorden die ongerelateerd zijn aan die taal. Verder onderzoek is nodig om te bepalen of het bekend zijn met de woorden voorafgaand aan het leren van de regels tussen de woorden een effect heeft – en of dat effect positief of negatief is.

Een laatste hoofdlijn van dit proefschriftonderzoek richt zich op het vermogen om te generaliseren. Tijdens de experimenten met volwassen deelnemers heb ik kunnen laten zien dat gevoeligheid voor *a_b*-afhankelijkheden zich uitbreidt naar het vermogen om deze te herkennen in nieuwe *aX'b*-strings, waar *X'* nog niet eerder is tegengekomen. Dit vermogen om afhankelijkheden te generaliseren naar nieuwe contexten is belangrijk als NADL gebruikt moet worden om volledig productieve morfosyntactische regels te verwerven. Daarom heb ik, gebruikmakend van dezelfde procedure als voorheen, getest of kinderen van 18 maanden oud ook in staat zijn om *a_b*- afhankelijkheden te generaliseren naar nieuwe *aX'b*-contexten. De resultaten tonen niet aan dat kinderen onderscheid kunnen maken tussen correcte $a_iX'b_i$ en incorrecte $a_iX'b_j$ reeksen. Het zou daarom betekenen dat kinderen van 18 maanden oud alleen *a_b*-regels kunnen volgen in bekende contexten. Toekomstig onderzoek moet uitwijzen of kinderen dit vermogen om te generaliseren op een later moment ontwikkelen.

Zoals gesteld, in dit proefschrift heb ik evidentie proberen te vinden voor de stelling dat NADL een betrouwbaar mechanisme is voor het verwerven van morfosyntactische afhankelijkheden in natuurlijke talen. Ik heb laten zien dat de prosodische eigenschappen van functionele morfemen in natuurlijke talen behulpzaam zijn bij het ontdekken van afhankelijkheden tussen deze elementen. Ik heb daarbij onderzocht of aanvankelijke bekendheid met individuele functionele morfemen de latere verwerving van afhankelijkheden tussen deze morfemen zou kunnen ondersteunen. Tot slot heb ik evidentie gevonden voor de stelling dat kinderen van 18 maanden oud in staat zijn om afhankelijkheden tussen prosodisch ongemarkeerde morfemen op te sporen over materiaal heen dat perceptueel meer saillant is. Ik heb echter geen evidentie gevonden voor de stelling dat kinderen van dezelfde leeftijd de afhankelijkheden die ze hebben geleerd ook kunnen generaliseren naar nieuwe contexten. Dit onderzoek ondersteunt daarmee de hypothese dat NADL als een geschikt mechanisme kan worden gezien voor taalverwerving, en dat dit nieuwe deuren opent voor verder onderzoek van deze hypothese.

Curriculum Vitae

Ileana Grama was born on September 15, 1987, in Bucharest, Romania. She studied Foreign Languages and Literatures (with a major in English and a Minor in French) at the University of Bucharest between 2006 and 2009, graduating with a Bachelor's degree. In 2009 she started a Research Master's program at Utrecht University, entitled *Linguistics: The Study of the Language Faculty*. She majored in Experimental Psycholinguistics and minored in Syntax. She graduated *cum laudae* in 2011, when she was awarded an NWO grant for an individual PhD project at the Utrecht Institute of Linguistics OTS. This dissertation is the result of her PhD research. Since 2015, Ileana has been teaching at Utrecht University and University College Utrecht.