UNIVERSITY OF QUEBEC IN MONTREAL

ABSTRACTION OF RULES AND THE LEARNING OF EXCEPTIONS IN FOURTEEN-MONTH-OLD INFANTS

THESIS

PRESENTED

AS A PARTIAL REQUIREMENT

OF A DOCTORATE IN PSYCHOLOGY

BY

ELENA KOULAGUINA

JANUARY 2014

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ABSTRACTION DE RÈGLES ET APPRENTISSAGE D'EXCEPTIONS CHEZ LES ENFANTS DE QUATORZE MOIS

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN PSYCHOLOGIE

PAR

ELENA KOULAGUINA

JANVIER 2014

REMERCIEMENTS

En premier lieu, je tiens à remercier ma directrice de thèse, Prof. Rushen Shi, pour toutes ses qualités de grande chercheure qu'elle investit sans cesse dans ses étudiants avec beaucoup de passion pour son travail. Sans elle, ce travail n'aurait jamais été possible.

J'exprime mes remerciements les plus profonds à ma famille et surtout à ma mère, Valentina Koulaguina.

Je souhaiterais aussi exprimer ma grande gratitude à Etienne Harnad, Yuriko Oshima-Takane, Francisco Goire, André Achim, Marie Labelle et Jean Bégin pour leurs précieux commentaires et suggestions.

Je tiens à remercier Alexandra Marquis, Mireille Babineau, Andréane Melançon, Erin Robertson, Cristina Name, Sarah Massicotte-Laforge, Camille Bédard, Sarah-Jeanne Viau, Laurie-Anne Dion, Alexandre Michaud et Jun Gao pour leurs diverses contributions.

Je remercie surtout les enfants qui ont participé à ces études et leurs parents.

TABLE OF CONTENTS

Page

LIST OF FIGURES viii

LIST OF ABBREVIATIONS xi

RÉSUMÉ xii

SUMMARY xv

INTRODUCTION 1

CHAPTER I
LITERATURE REVIEW AND RESEARCH QUESTIONS 5

1.1 Rule learning and generalization in infants 6

1.1.1 Pattern learning and rule generalization in infants 6

1.1.2 Learning of adjacent grammatical relations 12

1.1.3 Learning of non-adjacent grammatical relations 24

1.1.4 Infants' generalization when two rules are possible 34

1.2 Generalization, regularization and over-generalization 37

1.2.1 Defining the terms of generalization, regularization and over-generalization 37

1.2.2 Generalization studies 40

1.2.3 Regularization studies 41

1.2.4 Over-generalization errors in children's productions 49

1.3 Models of exception learning 52

1.4 Type and token frequencies in abstraction of rules and learning of specific examples 60

1.5 Infants' sensitivity to morphological markings 65

1.6 Brief summary of the literature review 68

1.7 Research questions and hypotheses 69

CHAPTER II
EXPERIMENTS 72

2.1 Rule learning and generalization from 100% consistent input 72
2.1.1 Experiment 1: Training – 100% consistent; test – novel instances; no morphological markings; age – 11-month-olds 72
2.1.2 Experiment 2: Training – 100% consistent; test – novel instances; no morphological markings; age – 14-month-olds....................79
2.2 The role of type frequencies in infants' learning and generalization of movement rules 81
2.2.1 Experiment 3: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; morphologically marked 81
2.2.2 Experiment 4: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; morphologically marked 88
2.2.3 Experiment 5: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; no morphological markings 91
2.2.4 Experiment 6: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; no morphological markings 96
2.3 Rule generalization and the role of morphological markings when the level of noise is high and the exposure to the input is increased 99
2.3.1 Experiment 7: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; morphologically marked; increased training 99
2.3.2 Experiment 8: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; no morphological markings; increased training 101
2.3.3 Experiment 9: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; training morphologically marked, test unmarked; increased training 104
2.3.4 Experiment 10: Training – 50% types of rule instances; type-token ratio of rule sentences 1:4, type-token ratio of noise 1:4; test – novel instances; training partially marked, test unmarked; increased training 106
2.4 The nature of rule encoding 109
2.4.1 Experiment 11: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; training morphologically marked, test unmarked 109
2.4.2 Experiment 12: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; training partially marked, test unmarked 112
2.5 Over-generalization and learning of exceptions 115

2.5.1 Experiment 13: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – noise instances from the training; morphological markings 115
2.5.2 Experiment 14: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – noise instances from the training; morphologically marked 119
2.5.3 Experiment 15: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – noise instances from the training; no morphological markings 122

DISCUSSION AND CONCLUSION 125

APPENDIX A
SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 1 AND 2 148

APPENDIX B
SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENTS 1 AND 2 149

APPENDIX C
SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENTS 3, 4 AND 7 150

APPENDIX D
RULE-CONFORMING SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 3, 4, 7, 9, 11, 13 AND 14 151

APPENDIX E
NOISE SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 3, 11, 12, 13 AND 14 152

APPENDIX F
NOISE SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 4, 7, 9 AND 10 153

APPENDIX G
SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENTS 5, 6, 8, 9, 10, 11 AND 12 154

APPENDIX H
RULE-CONFORMING SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 5, 6, 8 AND 15 155

APPENDIX I
NOISE SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 5 AND 15 156

APPENDIX J
NOISE SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 6 AND 8 157

APPENDIX K
RULE-CONFORMING SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 10 AND 12 158

APPENDIX L
SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENTS 13 AND 14 159

APPENDIX M
SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENT 15 160

APPENDIX N
INFANTS' EXPOSURE TO LANGUAGES ACCORDING TO PARENTAL REPORT 161

REFERENCES 164

LIST OF FIGURES

Figure Page

2.1 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 1: Training – 100% consistent; test – novel instances; no morphological markings; age – 11-month-olds. Infants' looking times for the two types of test trials were not significantly different 78

2.2 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 2: Training – 100% consistent; test – novel instances; no morphological markings; age – 14-month-olds. Infants' looking times for the two types of test trials were significantly different 80

2.3 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 3: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; morphologically marked. Infants' looking times for the two types of test trials were significantly different...87

2.4 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 4: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; morphologically marked. Infants' looking times for the two types of test trials were not significantly different 90

2.5 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 5: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; no morphological markings. Infants' looking times for the two types of test trials were significantly different 95

2.6 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 6: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; no morphological markings. Infants' looking times for the two types of test trials were not significantly different 98

2.7 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 7: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; morphologically marked; increased training. Infants' looking times for the two types of test trials were significantly different 100

2.8 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 8: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; no morphological markings; increased training. Infants' looking times for the two types of test trials were not significantly different 103

2.9 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 9: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; training morphologically marked, test unmarked; increased training. Infants' looking times for the two types of test trials were not significantly different 105

2.10 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 10: Training – 50% types of rule instances; type-token ratio of rule sentences 1:4, type-token ratio of noise 1:4; test – novel instances; training partially marked, test unmarked; increased training. Infants' looking times for the two types of test trials were not significantly different 108

2.11 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 11: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; training morphologically marked, test unmarked. Infants' looking times for the two types of test trials were not significantly different 111

2.12 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 12: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; training partially marked, test unmarked. Infants' looking times for the two types of test trials were significantly different 114

2.13 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 13: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – noise instances from the training; morphologically marked. Infants' looking times for the two types of test trials were not significantly different 118

2.14 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 14: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – noise instances from the training; morphologically marked. Infants' looking times for the two types of test trials were significantly different 121

2.15 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 15: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – noise instances from the training; no morphological markings. Infants' looking times for the two types of test trials were not significantly different 124

LIST OF ABBREVIATIONS

ACC	Accusative
CAUS	Causative
CV	Consonant – Vowel
e.g.	for example
et al.	and others
i.e.	that is
ISI	interstimulus interval
M	Mean
NP	Noun Phrase
SD	Standard Deviation
SE	Standard Error
SOV	Subject-Verb-Object
SVO	Subject-Object-Verb
VAP	Verb-Agent-Patient
VP	Verb Phrase
VPA	Verb-Patient-Agent

RÉSUMÉ

Cette thèse examine la généralisation des règles et l'apprentissage d'exceptions chez des jeunes enfants. Premièrement, notre travail cherche à expliquer comment les jeunes enfants arrivent à apprendre des règles abstraites à partir d'exemples concrets et comment ils généralisent les règles à de nouveaux cas. Deuxièmement, nous tenterons de mieux comprendre la façon dont les enfants traitent des cas ne conformant pas à une règle : sont-ils enclins à faire une sur-généralisation de la règle et de l'appliquer à des cas auxquels elle ne s'applique pas, ou apprennent-ils chaque cas comme si celui-ci constituait une exception? Nous adressons également la question du type de généralisation faite par les enfants lorsque deux règles sont possibles à partir des mêmes stimuli. En particulier, notre objectif est de comprendre le rôle des fréquences de types et d'occurrence dans ces processus d'apprentissage.

Une recherche antérieure a montré que les jeunes enfants sont capables de généraliser une règle abstraite à de nouveaux cas, malgré la présence d'items qui violent ladite règle dans les stimuli qui leur sont présentés (Gómez et LaKusta, 2004). Les généralisations ont échoué cependant quand la quantité de bruit (c'est-à-dire, les stimuli ne conformant à la règle) fut augmentée par fréquence de type et par fréquence totale. Aucune recherche n'avait encore testé le rôle de la fréquence de type dans la généralisation chez les jeunes enfants. Nous avons émis une hypothèse voulant qu'une prépondérance par fréquence de type de cas conformant à la règle, relativement aux cas de bruit n'y conformant pas, permettrait la généralisation des règles abstraites.

Une étude par Gerken (2006) a examiné la généralisation chez des enfants lorsque deux règles abstraites étaient identifiables à partir des mêmes stimuli. La recherche démontra qu'en présence d'un tel dilemme les enfants choisissaient d'interpréter les stimuli selon la règle la plus conservatrice. Continuant dans la lignée établie par Gerken (2006), ce travail tente d'identifier quelle généralisation est choisie par les enfants quand les stimuli d'apprentissage contiennent des cas de bruit. Nous avons émis l'hypothèse que les enfants procéderaient une généralisation conservatrice, de façon similaire à Gerken (2006).

D'autres études avec des enfants et des adultes ont démontré l'impact de la fréquence d'occurrence pour l'ancrage des verbes dans les structures de grammaire (Brooks et al., 1999, Theakston, 2004; Ambridge et al., 2007; Wonnacott, Newport et Tanenhaus, 2008). Ceux-ci mirent en évidence qu'enfants et adultes faisaient plus d'erreurs de sur-généralisation avec des verbes de basse fréquence (Brooks et al., 1999, Theakston, 2004; Ambridge et al., 2007). Nous avons émis l'hypothèse qu'une

haute fréquence d'occurrence des cas de bruits mènerait à l'apprentissage d'exceptions, tandis que leur basse fréquence d'occurrence mènerait à la sur-généralisation.

Dans le but d'examiner la généralisation et l'apprentissage des exceptions chez des enfants, nous avons construit deux règles artificielles de mouvement de l'ordre de mots (ABC – BAC et ABC – ACB), utilisant la langue russe, car celle-ci était entièrement étrangère à nos participants (des enfants âgés de 11 et de 14 mois). Certaines expériences incluaient aussi quelques cas de bruit. Les phrases avec du bruit étaient d'un type qui n'était pas sujet à aucun mouvement dans l'ordre des mots (ABC).

Pendant l'entraînement, les enfants écoutaient un nombre de phrases qui conformaient à l'une des règles. Dans certaines expériences, les phrases qui suivaient la règle étaient mélangées avec des phrases présentant le bruit. Après l'entraînement, les enfants furent testés selon une procédure de fixation centrale du regard. Pendant le test, les enfants écoutaient deux types d'essais : les phrases qui conformaient à la règle qu'ils avaient rencontrée lors de l'entraînement et les phrases qui conformaient à la règle pour laquelle ils n'étaient pas entraînés. Nous mesurions ici le temps de fixation de leur regard pendant l'écoute de ces deux types d'essais du test. Dans la phase de test des expériences sur la généralisation, les règles étaient appliquées à des nouvelles phrases que les enfants n'avaient jamais entendues auparavant. Dans la phase de test des expériences sur l'apprentissage d'exceptions, les règles étaient appliquées aux phrases de bruit qui n'étaient pas sujets à un mouvement dans l'ordre des mots pendant la phase d'entraînement.

À travers une série d'expériences, nous avons manipulé la fréquence de type et la fréquence d'occurrence des phrases qui présentaient la règle et le bruit dans l'entraînement. Nous avons aussi manipulé la régularité des marqueurs morphologiques dans l'entraînement et dans le test.

Les résultats ont démontré que les fréquences de type et d'occurrence ont eu un impact sur la généralisation et l'apprentissage des exceptions chez des jeunes enfants. Quand la fréquence de type des cas respectant la règle fut proportionnellement haute dans l'entraînement, la généralisation de la règle à des nouveaux cas a eu lieu. La généralisation n'a pas eu lieu quand les cas de règle et les cas de bruit étaient égaux en terme de fréquence de type lors de l'entraînement. Nous avons trouvé par ailleurs que des enfants font une généralisation plus conservatrice lorsqu'est introduite dans les stimuli une ambiguïté permettant deux niveaux de généralisations, un niveau plus large (plus abstrait) et un niveau plus étroit. Nos résultats démontrent également que la fréquence d'occurrence des cas de bruit peut affecter la sur-généralisation et l'apprentissage des exceptions. Une basse fréquence d'occurrence de cas de bruit a mené à la sur-généralisation d'une règle abstraite, tandis que la haute fréquence

d'occurrence de bruit a mené à l'apprentissage d'exceptions (c'est-à-dire, une résistance à la sur-généralisation).

MOTS CLÉS : généralisation des règles, apprentissage d'exceptions, enfants, fréquence de type, fréquence d'occurence

SUMMARY

This thesis examines rule generalization and the learning of exceptions in infants. One question addressed by our work is how infants learn abstract rules from concrete exemplars and generalize the rules to novel instances. Another question is how infants treat instances that do not conform to the rule: do they over-generalize them to a more general rule or learn them as exceptions? We also address the question of what kind of generalization infants make when two rules are possible based on the same input. More specifically, our goal is to understand the role of type and token frequencies in these learning processes.

Previous research showed that infants can generalize abstract rules to novel instances despite the presence of some rule violations in the input (Gómez and LaKusta, 2004). However, generalization failed when noise utterances increased in type and overall frequency. The role of rule type frequency in generalization by infants has not been specifically tested in previous research. We predicted that a high frequency of rule types relative to the noise promotes the generalization of abstract rules.

Gerken (2006) reported that when two rule interpretations are possible, infants choose the more conservative interpretation. We extended this work to the question of what generalization infants choose when the learning input supporting two interpretations contains some noise. We hypothesized that infants' generalization would be conservative, similarly to the study by Gerken (2006).

Other studies with children and adults demonstrated the impact of token frequency for the entrenchment of verbs in particular grammatical structures (e.g., Brooks et al., 1999; Theakston, 2004; Ambridge et al., 2007). It was found that children and adults make more over-generalization errors with low frequency verbs. We predicted that a high token frequency of noise exemplars would lead to the learning of exceptions, while low token frequency would lead to over-generalization of noise instances to the rule.

To examine infants' generalization and learning of exceptions, we constructed two artificial word order movement rules (ABC –> BAC and ABC –> ACB) using Russian, a natural language unknown to our infant participants (aged 11 and 14 months). For example, a sentence with ABC word order was followed immediately by the same sentence transformed into BAC in one movement rule (ABC –> BAC), or into ACB in another movement rule (ABC –> ACB). Some experiments also contained noise instances. The noise sentences did not undergo any movement (ABC).

In the training, infants heard a number of sentences conforming to one of the two rules. In some experiments, the rule-based sentences were intermixed with noise sentences. After the training, infants were tested with a central fixation preference procedure. In the test, infants listened to two types of sentences: some conforming to the trained rule and some conforming to the non-trained rule. The measure was the time spent looking towards the screen while listening to the two kinds of sentences. In the test phase of the experiments on generalization, the rules were applied to novel sentences that the infants had never heard before. In the test phase of the experiments on the learning of exceptions, the rules were applied to noise sentences that did not have movement during the training phase.

Across the series of experiments, we manipulated the type and token frequencies of rule and noise sentences during the training phase. We also manipulated the consistency of morphological markings in the training and test phases.

We found that type and token frequencies influence rule generalization and the learning of exceptions in infants. A high type frequency of rule instances in the training phase promoted generalization to novel instances. Generalization was impeded when rule and noise instances in the training phase were equal in type frequency. Infants also made more conservative generalizations when the examples allowed two rule generalizations, one bigger (more abstract) and another smaller. The smaller, more conservative generalization was more closely tied to the properties of the input: the presence of morphological markings. It was also found that the token frequency of noise exemplars can affect over-generalization and the learning of exceptions. Low token frequency of noise exemplars favored over-generalization to the abstract rule, while high token frequency of noise exemplars favored the learning of exceptions (i.e., resistance to over-generalization).

Key words: generalization of rules, learning of exceptions, infants, type frequency, token frequency

INTRODUCTION

This thesis addresses three questions concerning learning in infants. The first is about how infants learn abstract rules from concrete exemplars and generalize the rules to novel instances. The second is about whether infants make over-generalizations of instances that do not conform to the rule and how exceptions are learned. The third about what kind of generalization infants make when two rules are possible.

Rule learning is a crucial part of human cognitive capacity. Abstract rules are part of human knowledge. They are often learnt from concrete exemplars. For example, if learners encounter a sequence of such items as *le le di*, *wi wi je*, *de de we* etc., they can abstract a general rule AAB. This rule can be extended to other, novel, instances (e.g., *ba ba po*, *ko ko ga*). The ability to abstract a rule and generalize it to novel instances was recently shown in infants under one year of age (Marcus et al., 1999). This capacity is fundamental for language acquisition since language also involves rules, although syntactic rules are more complex. Syntactic rules are also learned by children from concrete exemplars. In the acquisition of their first language, children are never taught syntactic rules explicitly. This is different from second language acquisition in which rules are usually formally explained to learners.

Previous research has shown that rule generalization can tolerate a certain level of noise, i.e. instances not conforming to the rule. In a study by Gómez and LaKusta (2004), infants' generalization was successful when the training input contained 20 rule-conforming instances and 4 rule-violating exemplars. However, infants' generalization failed when the training contained 16 rule-conforming instances and 8 rule-violating exemplars.

One important question about rule abstraction is what children do when more than one generalization is possible. Such training input was presented to infants in a study by Gerken (2006). Infants were trained with a sequence of items such as *le le di*, *wi wi di*, *ji ji di*, *de de di*. This input is compatible with two rules. A more general rule is AAX, where both A and X positional categories can be generalized to novel items. A more conservative rule is AA*di*, where novel items can be used only in the A positions and the *di* item is always required in the final position. Trained under this condition, infants chose the more conservative rule.

In the case where an abstract rule can be formed despite some noise, those non-rule-conforming exemplars are exceptions to the general rule. Exceptions are very common in natural languages. It is known that young children over-generalize rules despite having heard exceptions (e.g., Bowerman, 1988). For example, English-learning children often produce *go-ed* instead of *went*. What are the factors underlying over-generalization? How do children eventually learn exceptions? Previous research with adults (Wonnacott, Newport and Tanenhaus, 2008) suggests that high-frequency verbs are learned in the specific grammatical structure where they tend to occur. It was also shown that children and adults make more over-generalization errors with low-frequency verbs (Brooks et al., 1999, Theakston, 2004; Ambridge et al., 2008). The role of the frequency of the kinds of examples encountered in over-generalization and exception learning has not yet been tested specifically with very young infants.

The goal of this work is to examine the role of type and token frequencies in the generalization of abstract rules and the learning of exceptions. Type means a kind of thing. In our case, individual grouping of words in a sentence makes one sentence type. Token means the number of occurrences of a type. In our case, each occurrence of a sentence type makes one token. We sought to replicate some of the results described above and to extend them to new findings with infants. In Experiments 1 and 2 we aimed at replicating the results of Marcus and colleagues (1999). Rule generalization was tested in the ideal learning situation where the training input did not contain any noise. We constructed an artificial word order movement rule (e.g., ABC –> BAC) using Russian, a natural language unknown to our infant participants. For example, a sentence with ABC word order was followed immediately by the same sentence transformed into BAC in one movement rule (ABC –> BAC; e.g., *Veter vybil okna –> Vybil veter okna*), or into ACB in another movement rule (ABC –> ACB; e.g. *Veter vybil okna –> Veter okna vybil*).

In Experiments 3 – 10, the training input contained noise sentences which did not follow the rule (e.g., ABC). We manipulated the type frequency, i.e., the number of different sentence types representing the noise. We also manipulated the number of occurrences for each noise sentence, i.e., the token frequency. These manipulations changed the relative frequencies of types and tokens of rules and noise in the training. We tested the effect of these different frequencies on rule generalization.

Experiments 11 and 12 tested the nature of rule learning when more than two generalizations were possible: a larger generalization and a smaller generalization. We examined which input properties would lead infants to one generalization or another. The goal was to test whether in the presence of noise infants would make a generalization that fit more conservatively with the input properties, as they did in (Gerken, 2006), who used input without any noise.

Experiments 13 – 15 examined the conditions under which infants over-generalize or learn exceptions. These experiments served to extend the studies with adults (Wonnacott, Newport and Tanenhaus, 2008) and children (Brooks et al., 1999, Theakston, 2004; Ambridge et al., 2008) to infants. We tested whether low token frequency of noise exemplars would lead to over-generalization, and whether high token frequency of noise would lead to the learning of exceptions.

CHAPTER I

LITERATURE REVIEW AND RESEARCH QUESTIONS

In their first two years, infants face a challenge of abstracting and learning various syntactic regularities, even when semantic cues are incomplete or absent. The first section of this literature review is focused on rule learning and generalization in infants. It includes a review of studies on learning and generalization of abstract patterns, adjacent and non-adjacent grammatical relations. It also covers the known research on infants' generalization when two interpretations of a rule are possible. The second section describes studies that provide evidence of learners sometimes going beyond the statistical properties of the specific input. Thus, when children learn a grammatical rule from the inconsistent input, they sometimes apply it to all novel instances, showing over-generalization. This tendency of children to over-generalize raises a question – how do children learn that some particular cases are exceptions, i.e. not subject to generalization? This question was approached in theoretical models of exception learning that are reviewed in the third section. Given that infants in the beginning of their language acquisition have no knowledge of semantic information, a matter of interest in their learning situation is the model of entrenchment, where specific words are entrenched to a grammatical structure if they occur in the input frequently. The fourth section examines studies that show infants' sensitivity to morphological markings. The fifth section gives a brief summary of the literature review. The sixth section covers research questions and hypotheses of the present work.

1.1 Rule learning and generalization in infants

1.1.1 Pattern learning and rule generalization in infants

The present section reviews studies on pattern learning and rule generalization in infants. Particular attention is paid to research showing learners going further than the mere tracking of specific co-occurring elements to attain a demonstrated capacity for rule abstraction and generalization to novel instances.

An important study was conducted by Marcus and colleagues (1999). Using simple identity patterns (i.e., the same element being repeated), they examined whether preverbal infants could learn abstract relationships and generalize them to novel instances. Participants were seven-month-old infants. In Experiment 1, one group of infants was familiarized with the ABA pattern (e. g., *ga ti ga*; *li na li*), whereas another group of infants was familiarized with the ABB pattern (e.g., *ga ti ti*; *li na na*). The stimuli were presented in triplets. In the test phase, both groups of infants heard two types of test trials: one type conformed to the trained pattern (ABA for the first group of infants and ABB for the second group of infants), whereas another trial type conformed to the untrained pattern (ABB for the first group of infants and ABA for the second group of infants). Both patterns were applied to completely novel items which were not previously heard by infants during the familiarization phase (e.g., *wo fe wo* representing the ABA pattern versus *wo fe fe* standing for the ABB pattern). Infants were tested with the familiarization adaptation of the Head-turn Preference Procedure. In the familiarization phase, infants listened to stimuli playing simultaneously from two side speakers while a central light was flashing in front of them. In the test, either the right or the left light started flashing. Once an infant fixated on that light, a speaker from the same side started playing a test trial. Each test trial lasted till infants looked away for two seconds or till the maximal trial length was reached. In the test phase infants showed a preference for the untrained pattern. These results suggested that infants learned the abstract patterns during familiarization and generalized them to novel items in the test.

Items used by Marcus et al. (1999) had a syllabic consonant-vowel shape (e.g., *ga*, *ti*, *wo*). In Experiment 1, test stimuli used for the A and B categories had distinct phonetic profiles: items in the A category started with a voiced consonant, whereas B category items started with an unvoiced consonant. This voicing pattern occurred in 18.75% of familiarization utterances (i.e., in 3 out of a total of 16 utterances). Although the pattern was not consistent across the familiarization stimuli, the authors decided to control for the possibility that infants could pick up this voicing pattern in the familiarization and recognize it in the test stimuli. For that purpose, Marcus et al. (1999) conducted Experiment 2. The voicing in the test stimuli was no longer consistent: *ba po ba* and *ko ga ko* represented the ABA pattern whereas *ba po po* and *ko ga ga* represented the ABB pattern. The familiarization stimuli were made of items which all started with voiced consonants (e.g., *le di le, wi je wi* for ABA pattern, and *le di di*, *wi je je* for ABB pattern). Infants again discriminated between the trained and untrained patterns.

In Experiment 3, Marcus et al. (1999) balanced the immediate reduplication feature in two types of patterns. In the first two experiments, only ABB stimuli contained an immediately reduplicating item, whereas in another type of pattern, ABA, the immediate reduplication was absent. Hence, infants could solely pay attention to the presence or absence of the same item repeating twice without an intervening element. To control for that feature in Experiment 3, the authors replaced the ABA pattern by an AAB pattern, while keeping the design and stimuli items from the previous experiment. In Experiment 3, one group of infants was familiarized with the AAB pattern (e. g., *le le di*; *wi wi je*), whereas another group of infants was familiarized with the ABB pattern (e.g., *le di di*; *wi je je*). In the test, these two patterns were applied to novel items, identically to the design of the previous experiments. Again, infants showed a preference for the untrained pattern. These results suggested that infants' successful learning and generalization in the previous experiments was not merely the result of their discrimination of sequences containing an immediate reduplication of an item. The three experiments conducted by Marcus et al. (1999) suggested that seven-month-olds could learn and generalize simple rules at an abstract level.

This was the first study which showed that preverbal infants could learn simple identity-based patterns and generalize them to novel items. Rules used by Marcus et al. (1999) were based on the repetition of the same syllable: The ABA pattern differed from the AAB and ABB patterns by the presence of an intervening element between the repeated items; AAB and ABB patterns differed by the position of the immediately reduplicating items. Infants were able to generalize such patterns even when there was no consistent pattern of voiced and unvoiced consonants across the familiarization and test. Infants' generalization was equally successful when the reduplicated elements were adjacent (as in the AAB and ABB patterns) or non-adjacent (as in the ABA pattern). A subsequent study by Gerken (2006) replicated these results with nine-month-olds, but not with seven-month-olds.

Rule abstraction and generalization capacities found in infants were also observed in adults. For example, Wonnacott, Newport and Tanenhaus (2008) trained adults with a miniature artificial language containing two verb arguments structures (*Verb Agent Patient* and *Verb Patient Agent Particle* structures). One of the structures was dominant in the training. In the test, adults generalized the dominant structure to novel verbs. In a production task, they used novel verbs in a structure that had been dominant in the input. Similar results were found with English-speaking six-year-olds (Wonnacott, 2011) who were trained with artificial language sentences containing an

English noun followed by a nonce particle. The frequency of particle use was manipulated, while the structure of sentences was the same. One particle was more frequent in the input than another. Similarly to adults, children generalized novel nouns to the more frequent particle more often. More details of these studies (Wonnacott, Newport and Tanenhaus, 2008; Wonnacott, 2011) are discussed in Section 1.4.

A number of studies examined infants' abstraction of rules from non-linguistic stimuli. In all these studies, the training did not contain any noise, so they examined infants' abstraction of rules from 100% consistent input. Studies using visual stimuli showed that seven-month-olds could successfully abstract the rules when all the items of a pattern were presented to them simultaneously (Saffran, Pollak, Seibel and Shkolnik, 2007).

Johnson and colleagues (2009) examined the same capacity of learning and generalizing to novel examples of sequential visual patterns. Stimuli were visual items differing in color and shape (e.g., an orange triangle, a gray octagon, a red square, etc.). Items appeared in the centre of the screen in an order based on either an ABA, ABB, or AAB rule. For example, the ABA rule was presented in the following time sequences: octagon – square – octagon, chevron – diamond – chevron, etc. Some shapes were used in the training, whereas other, novel, shapes were kept for being shown in the test phase for the first time. Two groups of infants were tested: eight-month-olds and eleven-month-olds.

In some conditions, eleven-month-olds were familiarized with a pattern containing adjacent identity items. For one group of infants, it was an ABB familiarization pattern, and for two other groups of infants, it was an AAB familiarization pattern. In the test, infants were tested with novel sequences following a trained and an untrained pattern. The untrained pattern was either another adjacent identity pattern or a pattern with non-adjacent identity items, depending on conditions. More specifically, one group of infants was trained with ABB and tested with ABB versus AAB. Another group of infants was trained with AAB and tested with ABB versus AAB. The third group of infants was trained with AAB and tested with AAB ABA. In all three conditions, infants preferred AAB. This pattern was untrained for the first group of infants, hence, they showed a novelty preference in the test. For the other two groups of infants the pattern was trained; hence, they had a familiarity preference. A supplementary control group of infants was tested with AAB and ABB, without any preceding training. The control group did not show any preference in the test. The absence of discrimination in the control group suggested that results in the three experimental groups were related to infants' learning of training patterns. The total of experimental conditions showed that eleven-month-olds had a capacity of learning and generalizing adjacent identity relationships to novel stimuli.

Eight-month-olds showed mixed results. In some conditions, they did not show any discrimination. Thus, a group of infants trained with ABB did not discriminate it in the test from an untrained adjacent repetition pattern, AAB. A group of infants trained with AAB did not discriminate it from a nonadjacent ABA. However, a group of eight-month-olds trained with ABB showed a novelty preference for ABA in the test. To examine whether this was related to their processing of familiar stimuli, a supplementary control group was tested with ABB and ABA test sequences, without preliminary training. The control group did not show any preference. This suggested that in eight-month-olds, the capacity to learn and generalize from adjacent repetition patterns was only partially emerging. So far, eight-month-olds showed evidence of learning and generalizing such a pattern only when the repeated items occurred at the end of the pattern and when it was contrasted with a non-adjacent repetition pattern in the test.

In another experimental condition, eleven- and eight-month-old infants were trained with non-adjacent repetition (i.e., ABA) and tested with novel ABA versus untrained ABB. Both ages failed to show learning and generalization.

Studies using auditory stimuli showed that speech had an advantage for abstracting identity-based rules. Thus, Marcus, Fernandes & Johnson (2007) showed that infants could not abstract such rules as ABB, ABA or AAB from tones, sounds of different musical instruments (timbre) or animal calls. However, once they were familiarized with speech syllables (such as le di di, wi je je, ji li li etc.) they could discriminate trials with those abstract patterns even if in the test stimuli those patterns were composed of tones, sounds of musical instruments or animal calls. So, this suggests that at the level of extracting abstract rules, speech is privileged for infants over other non-speech auditory stimuli. However, abstraction could occur with visual stimuli.

The work reviewed above demonstrates that already during the first year of life infants have a strong capacity to detect consistent patterns based on the position of repeated elements in artificial languages (Marcus et al., 1999; Gerken, 2006; Johnson et al., 1999). Moreover, infants were able to apply this abstract knowledge to novel linguistic items (Marcus et al., 1999; Gerken, 2006) and novel visual and auditory stimuli (Johnson et al., 1999; Saffran et al., 2006; Marcus, Fernandes & Johnson, 2007). In the visual modality, patterns were better abstracted in simultaneously presented stimuli (Saffran, Pollak, Seibel and Shkolnik, 2007). When the items were presented sequentially, infants had more difficulty and could only generalize patterns with adjacent identity items (Johnson et al., 2009). This may be related to the fact that spatial processing is more characteristic of visual domain while in auditory domain temporal processing is more important. In the auditory modality, speech had an advantage over non-speech auditory stimuli: learning occurred for both adjacent and non-adjacent repetition when infants were trained with sequences of syllables

(Marcus et al., 1999; Gerken, 2006). Moreover, infants could generalize this knowledge to non-speech auditory stimuli (Marcus, Fernandes & Johnson, 2007). However, they could not abstract the patterns when the identity-based pattern in the training was made with non-speech stimuli such as tones, sounds of musical instruments or animal calls (Marcus, Fernandes & Johnson, 2007).

1.1.2 Learning of adjacent grammatical relations

This section reviews a number of studies on the learning of adjacent grammatical relations. Höhle and colleagues (2004) suggested that between 14 and 16 months of age, German-learning infants could use a German determiner *ein* to categorize the subsequent adjacent nonce words as nouns. One group of infants was familiarized with two target nonce words each preceded by the determiner. Another group was familiarized with those same target words, but preceded by the pronoun, thus using the targets as verbs. Afterwards, both groups were tested with two types of passages. In one, the nonce words from the training were used in the contexts requiring nouns, in the other, in the contexts requiring verbs. The determiner and pronoun were not used in the test passages. For one group of infants, the passages where nonce words were used in noun contexts were compatible with their training grammar, whereas the passages where the same words were used in verb contexts were incompatible with it. For the other group of infants, it was the opposite. Only the group familiarized with the determiner-noun sequences showed discrimination between two types of passages, with a novelty preference for passages where those words were used in verb contexts. The group of infants familiarized with pronoun-verb sequences did not show such discrimination. These results were interpreted as noun categorization, based on infants' knowledge of the German determiner *ein,* which preceded the target words in familiarization. The authors concluded that the infants' differential performance was linked to the distributional information in their learning input. To examine this hypothesis, they conducted a case study of spontaneous motherese, i.e., speech by a mother directed to her child during natural interactions. They found that by overall token frequency alone, the determiner *ein* was used about three times more frequently than the pronoun *sie*. Moreover, *ein* was a better predictor of nouns than *sie* was a predictor of verbs at the level of adjacent dependencies: *ein* was immediately followed by nouns in 71% of cases, whereas *sie* was immediately followed by verbs only in 31%. These results confirm that infants' different performance with nouns and verbs could be related to the different distributional information in their learning input.

Mintz (2006) also examined noun and verb categorization. English-learning one-year-olds were familiarized and tested with sequences where some nonce words were used in English noun frames and other nonce words were used in English verb frames. Although the test utterances contained the same nouns and verbs as the familiarization, they never presented the same specific combination of the target words and the adjacent frame. This design controlled for the infants' recognition of specific item combinations from familiarization. Here, unlike in the study by Höhle and colleagues (2004), infants showed a novelty preference for ungrammatical strings only for verbs, and not for nouns. This suggests that by one year of age, English-learning infants can use English verb frames, and not noun frames, to categorize nonce words. Although English-learning infants showed a different categorization pattern from German-learning infants, in both cases their performance appeared to be linked to the distributional information in the input. In English, verb frames are used more frequently than noun frames, as was shown in an analysis of a speech corpus by Mintz (2006), who suggested that this could result in better categorization for verbs than nouns in English-learning infants.

Shi and Melançon (2010) studied syntactic categorization in 14-month-old French-learning infants. Like Höhle et al. (2004), they examined noun and verb categorization based on the co-occurrence of target nonce words with functional words. French-learning infants were familiarized with two nonce words preceded either by pronouns (for one group of infants, in the pronoun+verb condition) or by determiners (for another group of infants, in the determiner+noun condition). In the test phase, both groups heard two types of test trials. In one type, the nonce words were preceded by a French pronoun that had never been presented during the familiarization. In another, the nonce words were preceded by a French determiner that had not been used in the training. Hence, the exact adjacent combination of a pronoun/determiner with a nonce word in the test was never encountered in the training. For each group, one of the test trial types was grammatical, i.e., the nonce word in the test was used with a function word consistent with its syntactic category in the training. Another test trial type was ungrammatical, i.e., the use of a function word was not compatible with the syntactic category of the nonce word. Infants showed the same patterns of results as German-learning infants: they were successful in noun categorization of nonce words preceded by determiners, but they failed in verb categorization based on the target word's co-occurrence with personal pronouns. Again, the distributional information in the language that infants were learning seemed to play a role. In French motherese, as in German, co-occurrence between determiners and nouns was more consistent than between pronouns and verbs.

A further step forward in research on infants' syntactic categorization is the work on infants' gender categorization of nouns based on their co-occurrence with gender-marked determiners in French (Cyr & Shi, 2013). French-learning infants were familiarized with nonce nouns preceded by French gender-marked indefinite determiners. Two of those nonce words followed a masculine determiner *un*; two other nonce nouns followed a feminine determiner *une*. In the test trials, infants heard the same nonce nouns preceded by gender-marked definite determiners, masculine *le* and feminine *la*. In one type of test trial, the combinations were grammatical, according to the gender of the determiners in the training. In another, the combinations were ungrammatical. Infants were thus tested with determiners which were never encountered during familiarization. Their discrimination would be based solely on their previous knowledge of the determiners and their corresponding gender categories. By twenty months of age infants showed evidence of discrimination between the grammatical and ungrammatical trials. These results suggest that by twenty months French-learning infants have abstracted gender contingency patterns and can categorize nonce nouns as feminine and masculine based on their co-occurrence with determiners.

The studies reviewed above show that infants abstract adjacent syntactic category patterns in the language they acquire. However, it is not completely clear what linguistic factors guided infants in their initial learning. Mintz (2006) proposed that frequent frames could influence infants' syntactic categorization, whereas Höhle et al. (2004) suggested that learning was done by bigrams containing content words with frequent determiners. The two ideas are similar in that learning depends on frequent anchor points. In the case of the bigrams, there is only one anchor point influencing the categorization of the adjacent element. In the case of frames, there are two anchor points, combined in a non-adjacent relation.

To determine what distributional factors guide learning, researchers have conducted experiments using languages unknown to the infants. Very precise manipulations with the training input can be made to assess learning. A number of studies examined learning processes in adults and children based on word co-occurrence patterns. Valian and Coulson (1988) showed that adults could categorize the words of the miniature artificial language into categories, defined by their co-occurrence with the preceding function-like words. Like in many natural languages, where each syntactic category can be preceded by specific functors, the miniature artificial language used by Valian and Coulson (1988) was composed of 'aA bB' and 'bB aA' sentences, where small letters corresponded to functors and big letters to content words. The content words in the language could belong to one of two categories: 'A' words, which could only be preceded by 'a' functors, and 'B' words, which could only be preceded by 'b' functors. To test the effects of relative functor and content word frequency, Valian and Coulson (1988) created two experimental conditions. In one of them, a group of participants was familiarized with a dialect with only two functors and twelve content words. Thus, functors were six times more frequent than content words (the high-frequency condition). In another condition, another group was familiarized with a dialect with four functors and six content words. Thus, functors were only 1.5 times more frequent than content words (the low-frequency condition). Both groups were tested with the same vocabulary, which either conformed to the structure of the artificial language or violated it, either at a surface level (i.e., swapping the positions of the functor and the content word, or presenting two content words in a row) or at a deeper level of syntactic categorization of the content word (i.e., combining content categories and functors wrongly). In grammatical test stimuli, the exact combinations of a functor and a content word were already presented during the training. Participants were asked to judge whether the new sentences were similar to the original ones or not.

Adults learned the high-frequency dialect at both surface and syntactic categorization levels. The learning of the low-frequency dialect, however, failed at the deeper syntactic categorization level, although the participants did successfully acquire the surface structure of the language. The learning of the low-frequency dialect improved only when the artificial language training was enriched with visual reference information, though it still remained lower than for the high-frequency. In the study by Valian & Coulson (1988), the participants were tested on the same vocabulary they had previously heard in the familiarization. Their knowledge of the correct combinations of functors and content words showed their learning of the content word categories. However, their performance could also have been influenced by their memory for specific combinations of functors and content words that had been presented during the training phase. The ungrammatical test phrases had never been present in the training.

In a subsequent study on syntactic categorization, Mintz (2002) also trained adults with an artificial language. Unlike Valian & Coulson (1988), he used frames: in a sequence of three words, the initial and the final words were in a non-adjacent frame relation. Another crucial difference was that Mintz's relevant test stimuli in were all novel combinations of the items from training. This controlled for any pure memorization of specific word combinations. Several words were used as the middle word across four non-adjacent dependencies. One of the middle words occurred with three out of four non-adjacent dependencies during training. In the test, it was inserted within that fourth non-adjacent dependency. During training, there was also a fifth non-adjacent dependency with different middle words that never occurred in the other dependencies. In addition, the training contained four distracting three-word utterances without any non-adjacent dependency. Two middle words from those distracting utterances also occurred within the fifth non-adjacent dependency.

Test stimuli included the utterance containing the middle word that did not occur with the fourth dependency during training, as well as the fifth non-adjacent dependency with the same middle word. Both test utterances were novel combinations of the middle word with non-adjacent dependencies from the training, i.e., the middle word never occurred with either of them. The use of the middle word in the fourth dependency from the training (the first utterance in the test) was category-conforming. The fourth non-adjacent dependency in the training contained middle words that also occurred within other dependencies where the middle word from the test also occurred. The use of the middle word in the fifth dependency from the training (the second utterance in the test) was not category-conforming. The fifth non-adjacent dependency in the training contained middle words that never shared any common non-adjacent dependency with the test middle word. (Apart from these two test trials of interest, the test stimuli also contained other utterances: the exact utterances from the training or novel combinations of words from the distracting utterances in the training).

Participants were asked to say whether test sentences were among those presented to them during the training, and the degree of their confidence for each response. To discriminate between category-conforming and non-conforming test utterances, participants needed to make a syntactic categorization: first, they needed to sort non-adjacent dependencies as supporting two different categories, on the basis of the middle words they contained; and second, they needed to categorize the middle word in the test as either grammatical or ungrammatical for each of these two categories. Even in such a complex generalization task, adults were sensitive to co-occurrence patterns of syntactic categories based on the distributional analysis of their training input. They discriminated between the grammatical and ungrammatical sentences. Their judgment was based on the correct categorization of the non-adjacent dependency and its middle element.

General pattern learning and rule abstraction in children and adults were also studied by Saffran (2001) in a study of the learning of predictive dependencies. In linguistic systems, there can be predictiveness between categories. In syntax, predictive dependencies concern the relationship between abstract categories such as the categorical dependency between determiners and nouns discussed so far. The study by Saffran (2001) examined the learning of predictive dependencies between abstract syntactic categories. Saffran (2001) adapted the artificial grammar used by Morgan and Newport (1981) to create sentences with a hierarchical phrase structure. Each sentence was composed of three phrases, two obligatory and one optional. Each phrase was a combination of word categories and a phrase, either obligatory or optional. The sub-hierarchical phrase, for its part, contained two word categories, one obligatory, and another optional. Various utterances were formed according to this grammar. Adults and seven-year-olds were exposed to this artificial language and subsequently tested with three forced-choice grammaticality judgment tests.

Two Rule Tests were based on the knowledge of certain rules of the artificial language (for example, if there is a word from one category, there should be a word from another category). Both children and adults showed knowledge of such abstract rules, with the experimental groups performing significantly better than chance and better than the control groups, which received no training prior to test. Both children and adults showed better performance for positive rules (i.e., *if… there must be…*) than for negative rules (i.e., *there cannot be…*).

In the Rule Tests, the words used in new sentences were exactly the same as the ones in the training stimuli, and their exact combinations in pairs and triplets were partially presented during the training. To control for participants' learning of those specific combinations, Saffran (2001) performed a supplementary analysis which showed that participants' grammaticality judgments were not influenced by the frequency of such combinations. For that analysis, each test item was coded according to the following criteria relative to the training input: the average frequency of pairs and triples within each utterance; a composite frequency for the initial and final word pairs and triples for each utterance; the number of novel word pairs which were never presented during the training; and the number of words by which each utterance "differed from the most similar sentence in the input" (Saffran, 2001, p. 502). Even when controlled for all these supplementary distributional factors, the effect of grammaticality remained the only significant predictor of the participants' performance. None of the covariates contributed significantly to the variance.

The third, Fragment Test, tested participants' sensitivity to predictive dependencies in terms of phrasal groupings, i.e., their knowledge of phrasal units and phrasal boundaries. Here, the 'grammatical' test stimuli were fragments of sentences which could be grammatically combined as phrasal units, whereas the 'ungrammatical' test stimuli were the fragments spanning across phrases (i.e., they occurred in the training language but did not conform to the pattern characteristic of the phrasal unit as a whole). Such 'grammatical' and 'ungrammatical' fragments can be compared to an English phrase *the dog* and a fragment spanning across the phrasal boundary *bit the*. Although the second fragment can be encountered in a grammatical speech corpus, it does not represent a grammatical phrasal grouping. Here, in the Fragment Test, both 'grammatical' and 'ungrammatical' test stimuli contained the exact pairs of words from the training. For that reason, the discrimination between the 'grammatical' and 'ungrammatical' test sentences could not be based on the learning of specific combinations, since both types of stimuli were balanced in that regard. In this test, a significantly better performance was observed in the experimental group than the control group in adults, whereas children in the experimental group displayed a marginally better performance than the control group. The performance of the experimental group was also better than chance. However, these results were mainly due to participants' sensitivity to the predictive dependencies. By 'violation of predictive dependencies' the authors meant a different level of predictiveness in phrasal units and in fragments spanning across a phrasal boundary. The across-boundary-fragments, although a part of a grammatical sentence, can violate predictive dependencies at a higher level of the grammar. For example, if the across-boundary-fragment contains two items of the neighboring phrases, but the first item is an optional item in the end of the first phrase, this item does not predict the occurrence of the following member of the second phrase. When the degree of violation of dependencies was included in the analysis as a covariate, the effect of the 'grammaticality' (i.e., sensitivity to phrasal units) completely disappeared in adults and only left a trend towards significance in children. These results suggest that participants relied strongly on predictive dependencies rather than on phrasal units. "The number of predictive dependencies violated by each test fragment was a stronger predictor of participants' judgments than whether or not a fragment was a phrase" (P504).

The study by Saffran (2001) showed that both children and adults were able to learn abstract relationships between categories based on predictive dependencies. The participants showed sensitivity to statistical properties of the artificial language in the absence of semantic information. Their learning situation resembled that of preverbal infants. However, the kind of grammar used by Saffran (2001) involved both adjacent and non-adjacent relations, as well as some other obligatory rules and optionality. For example, one of those rules did not involve either adjacent or non-adjacent relations: i.e., every sentence must contain at least one A-category word. The optionality involved an optional presence of a category within a phrase. Due to such optionality, the relations between the neighboring items were sometimes adjacent, and sometimes non-adjacent. Although adjacent and non-adjacent dependencies were optional in the grammar used by Saffran (2001), they were probably required to access the structure. Overall, the relations between categories in this grammar were very complex, and it was not clear from this study which was more crucial for the learner.

Another study which examined infants' artificial grammar learning and generalization is that by Gómez and Gerken (1999). In their experiment, the familiarization and test stimuli shared a common vocabulary. The grammar consisted of certain rules defining which particular word category could follow which category, allowing several word orders. The artificial words from familiarization were re-combined in the test, either following or violating the artificial grammar in endpoints or in internal combinations. One-year-olds showed discrimination between the grammatical and ungrammatical test stimuli after a short exposure to a miniature artificial language. In another experiment, the familiarized vocabulary was replaced by the novel vocabulary in the test, in both grammatical and ungrammatical sequences. In the test phase, infants distinguished-novel items, which were combined according to the same grammatical rules as the training strings, from novel items, which were arranged ungrammatically. However, it is not clear how and what exactly infants learned, since the training grammar presented combinations of categories each including multiple words, and novel items contained no cues that might help infants to relate them to previously heard words and their categories. It is possible that infants applied patterns of transitional probabilities between elements from the training set to the novel vocabulary, since grammatical and ungrammatical test sets had different levels of transitional probabilities. On the other hand, the transitional probabilities of the novel vocabulary could not be easily accessed, since the test procedure was fully infant-controlled, with novel sequences interrupted at variable moments.

Using another artificial language Gómez and LaKusta (2004) examined whether one-year-olds would categorize words as different classes by linking their structure (number of syllables) with adjacent words. Those adjacent words mimicked functors in natural languages. Like functional words, they were few but frequently encountered. They predicted the occurrence of the following 'content' word, defining whether it had to come from one or another class. For example, words *alt* and *ush* predicted that the following adjacent word had to be disyllabic, whereas *ong* and *erd* predicted the occurrence of a monosyllabic word. This learning situation simulated infants' grammatical categorization of content words using function words. One year-olds appeared to be successful in learning such regularities and even in extending the rule to novel disyllabic and monosyllabic words. As in the study by Marcus et al. (1999), infants learned linguistic rules and generalized them to novel instances. In the study by Gómez and LaKusta (2004), however, learning and generalization were more complex, based not simply on identifying relations between reduplicated elements (ABB vs. ABA), but on tracking grammatical regularities for two structure-based categories (disyllabic versus monosyllabic words).

Gerken, Wilson and Lewis (2005) also examined infants' syntactic categorization. Unlike Gómez and LaKusta (2004), who tested their infants with an entirely artificial language, Gerken, Wilson and Lewis (2005) used a natural language unknown to infants. They trained and tested American-learning 17-month-old infants with a partial Russian gender paradigm. The training consisted of Russian words of two genders, masculine and feminine. Each of the categories was marked by two inflections (e.g., all feminine words in the training were marked with *oj* and *u* inflections, whereas all masculine words were marked with *ya* and *yem* inflections). Four words in each category were withheld for the test. Those words were only heard by infants with one of the inflections, but not the other. For example, *vannoj* was never presented to infants during the training, although they heard its root with the other feminine inflection, i.e. *vannu*. Another feminine noun, *korovu*, was never presented during the training, and the only form of this word that children heard in the training was *korovoj*. Gerken, Wilson and Lewis (2005) examined whether infants could learn which endings were characteristic for each category of nouns and whether they could generalize this knowledge to words which they only previously heard with one of the inflections. Interestingly, infants were able to learn and generalize the gender paradigm, but only when 60% of familiarization items were double-marked: apart from the inflections, they had a derivational morpheme *–tel* for masculine and *–k* for feminine nouns. Only for such input could infants derive the grammatical endings for the words they never heard before. Crucially, they did not require the test stimuli to be double-marked with the same derivational morphemes (i.e., *–tel* and *–k*). The generalization was successful even to words that did not have those morphemes. The consistency of double-markings in the training stimuli appeared to affect infants' learning and generalization. When 60% of familiarization items were double-marked, they could learn the paradigm and extend their knowledge to stimuli with single markings; when no familiarization items were double-marked, infants' learning and generalization failed.

Most of the studies reviewed above showed the learning of abstract categories. A number of studies showed infants' categorization of nonce words based on adjacent function words of a natural language they acquire (Höhle et al., 2004; Mintz, 2006; Shi and Melançon, 2010; Cyr & Shi, 2013). Valian and Coulson (1988) demonstrated the learning of adjacent categories in adults, using an artificial language. Saffran (2001) examined predictive dependencies at the categorical level in both adults and children. Infants' learning of abstract grammatical categories was also examined with artificial languages (Gómez and Gerken, 1999; Gómez and LaKusta, 2004) and with a natural language unknown to infants (Gerken, Wilson and Lewis, 2005).

1.1.3 Learning of non-adjacent grammatical relations

In addition to adjacent relationships, infants must learn non-adjacent relationships between elements. Non-adjacent dependencies were used in some of the studies reviewed above, although they were not the central research question in those studies. One type of training input in the studies by Marcus et al. (1999) and Gerken (2006) presented an abstract rule based on a non-adjacent dependency (the ABA rule). The two reduplicated items were in a non-adjacent relation (A_A). Infants in these studies learned this non-adjacent dependency. This section reviews studies that directly examined the learning of non-adjacent relations.

A study that specifically tested whether adults could track non-adjacent relations was conducted by Newport and Aslin (2004). They used artificial language speech streams where syllables were combined without pauses or distinct word boundaries. In one type of language, non-adjacent relations were introduced between whole syllables, and in another type, between consonants or between vowels. In the first type of language (with non-adjacent syllabic relations), artificial tri-syllabic sequences were designed. Syllables had a consonant-vowel (CV) structure. The first and the last syllables belonged to one of several frames. In such frames, the first item predicted the occurrence of the last item with a probability of 1.0. The middle syllable in tri-syllabic words was variable. Each of the middle syllables could occur within any of the five frames. This pattern can be represented as AXB, where A predicts B, while X is variable. In the second type of language (non-adjacent relations between consonants), tri-syllabic sequences contained non-adjacent consonants, with intervening variable vowels. Each vowel could follow any of the consonants at any of the positions where a vowel was allowed. Such a pattern can be presented as C1Vx-C2Vx-C3Vx, where C1 is the first consonant predicting the following consonants C2 and C3, while Vx is a variable vowel. Another type of language was designed with non-adjacent relations between vowels. A schematic pattern for such language would be CxV1-CxV2-CxV3, where V1 is the first vowel predicting the following vowels V2 and V3, while Cx is a variable consonant. Each consonant can precede any vowel.

After being familiarized with stimuli having non-adjacent relations, adults in the study by Newport & Aslin (2004) were tested on some of the familiarization stimuli that had non-adjacent dependencies, versus tri-syllable stimuli that violated the non-adjacent dependencies. The test utterances were formed by taking words from the training and combining them either according to or against non-adjacent dependency patterns. For example, for the syllabic non-adjacent relations, the grammatical sequence would be A_1XB_1, whereas an ungrammatical non-dependent sequence would be A_1XB_2. In another experimental condition with syllabic non-adjacent relations, the ungrammatical non-dependent sequence could be either BAX or XBA. For non-adjacent relations between consonants, the grammatical sequence would be C1Vx-C2Vx-C3Vx, and the ungrammatical non-dependent sequence would be (...C3Vx C1Vx-C2Vx...) or (...C2Vx-C3Vx C1Vx...). Similarly, in the case of the language with non-adjacent relations between vowels, the grammatical sequence would be CxV1-CxV2-CxV3, and the ungrammatical non-dependent sequence would be (...CxV3 CxV1-CxV2...) or (...CxV2-CxV3 CxV1...). Participants were asked to judge which of the sequences in each pair was more similar to the training input. Adults were able to track and recognize non-adjacent dependencies between consonants and between vowels, but not between syllables.

Another study, however, provided contradicting evidence suggesting that adults were able to track non-adjacent dependencies at the syllabic level (Peña et al., 2002). In this study, participants were French speakers, not English speakers as in the study by Newport and Aslin (2004). Here, adults were trained with a language similar to the one used by Newport and Aslin (2004). The first and the last syllable (within AXB sequences) were co-dependent with a 1.0 transitional probability, and the middle element was variable. In the first experiment, segmentation cues were absent between any elements (within and across AXB sequences: ...AXBAXBAXB...). Adults were tested with AXB dependent sequences and non-dependent sequences (e.g., XBA) from their training. They had a significant preference for dependent (e.g., AXB) over non-dependent stimuli (e.g., XBA).

In two other experiments by Peña et al. (2002), the participants were tested with dependency-violation stimuli (e.g., XBA) versus the non-adjacent dependent syllables containing novel intervening X syllables. Although X elements of the dependent sequences were presented in the training, they never occurred in the middle position between A and B. More precisely, in the training they were presented as As or Bs from other non-adjacent dependencies. Thus, the strings of dependent stimuli were novel. As for the non-dependent test stimuli, the combinations of their fragments were presented in the training. Participants were not able to discriminate between the two types of test trials in the absence of segmentation cues. However, when a brief pause of 25 ms was introduced in the training between all AXB sequences, they succeeded in discriminating the two types of test trials. The novelty of this study was that it tested not simply the learning and tracking of specific non-adjacent dependencies with familiar intervening elements, but also the generalization of those non-adjacent dependencies to intervening elements that never occurred in that position during training.

The studies with adult participants reviewed above show inconclusive evidence on whether adults can track syllabic non-adjacent dependencies from a stream of stimuli in the absence of any acoustic segmentation cues. Thus, English-speaking participants in the study by Newport and Aslin (2004) did not demonstrate such a capacity, whereas French-speaking participants in the study by Peña et al. (2002) were able to track non-adjacent dependencies at a syllabic level. Peña and colleagues (2002) showed that when subtle acoustic cues were introduced (i.e., brief pauses marked the boundaries of non-adjacent frames), adults were able to generalize the trained non-adjacent dependencies to trained items with a novel intervening element. The results obtained by Peña et al. (2002) agree with the findings on infants who were able to learn non-adjacent relations (e.g., Marcus et al., 1999). However, the stimuli used by Marcus et al. (1999) were simpler since non-adjacent frames were made with repeated items. Infants were able to generalize the relations between non-adjacent repeated items to novel repeated items (Marcus et al., 1999). The stimuli in Marcus et al. (1999) were made of uni-syllabic items with a CV structure. Whereas Peña et al. (2002) examined the generalization of an intervening element, Marcus et al. (1999) explored the generalization of the non-adjacent frame itself. Another difference is that non-adjacent relations in Peña et al. (2002) were combinations of several syllables, whereas in Marcus et al. (1999), they were reduplicated items. In both studies, the generalization was possible after learning from a training set that segmentation cues between sequences with a non-adjacent relation. In the study with adults, in the experimental condition where the generalization was successful, the pause was extremely brief, 25 ms. In another experimental condition, where the pause cue was absent, adults' generalization failed. Infants were trained only with stimuli where a salient pause of 1 s separated sequences of non-adjacent frames. Although the tasks in these two studies were not identical, they suggest that infants and adults are able to generalize patterns with non-adjacent dependencies to novel sequences after a short exposure to an artificial language.

Several studies examined infants' learning of non-adjacent dependencies in natural languages they acquire. Santelmann and Jusczyk (1998) examined whether English-learning infants could discriminate the grammatical non-adjacent relation of an auxiliary *is* with *–ing* morpheme (e.g., *At the bakery, everybody is baking bread)* from an ungrammatical non-adjacent relation between *can* and *–ing* morpheme (e.g., *At the bakery, everybody can baking bread)*. Middle elements were adverbs and verb roots. Adverbs varied by the number of syllables across experiments. At 18 months, but not at 15 months, infants were sensitive to the non-adjacent co-occurrence of the auxiliary *is* with the *–ing* morpheme. They could track this relation even when the intervening element consisted of three syllables, i.e., two-syllable adverb and single syllable verb root.

Interesting findings were obtained in a study which examined the tracking of German adjacent and non-adjacent dependencies by German-learning infants (Höhle et al, 2006). Nineteen-month-old infants were able to track an adjacent dependency between an auxiliary *haben* (*have*) and a past participle, forming a present perfect construction (e.g., *Das kleine unzufriedene Kind hat geheult – The little unhappy child has cried*). They preferred listening to passages with grammatical present perfect constructions, rather than to passages with ungrammatical combinations of an auxiliary *können* (*can*) followed by a past participle (e.g., *Das kleine unzufriedene Kind kann geheult – The little unhappy child can cried*). However, infants were not able to discriminate between grammatical present perfect constructions and ungrammatical utterances when the auxiliary and the participle were separated by an intervening adverb consisting of two syllables. These results were contradictory to those obtained by Santelmann and Jusczyk (1998) where eighteen-month-old English-learning infants were able to track the non-adjacent dependencies in an English progressive construction, with intervening elements up to three syllables (a two-syllable adverb and single syllable verb root). This difference of the two studies was surprising since in German non-adjacent elements of the present perfect construction typically allow more than one word as an intervening element, whereas for the English progressive construction, multiple intervening elements are not typical. To examine whether a syntactic category of the intervening element could play a role in infants' tracking of German present perfect construction, Höhle et al. (2006) tested infants with passages where, instead of an adverb, the intervening element was a complement, consisting of a determiner and a noun (e.g., *Das kleine phantasievolle Kind hat den Ball geholt – The little imaginative child has the ball fetched*). This time, infants preferred such grammatical constructions over the ungrammatical passages where an auxiliary *können* ('can') was wrongfully combined with a past participle over intervening elements (e.g., *Das kleine phantasievolle Kind kann den Ball geholt - The little imaginative child can the ball fetched*). These results suggest that infants' tracking of non-adjacent relations is sensitive to syntactic category of the intervening element. They were successful when the intervening element was a complement and failed when it was an adverb.

However, there is a possibility that in the study by Höhle et al. (2006), infants' successful tracking of non-adjacent dependencies separated by a complement was due to their tracking of frequent adjacent combinations of an auxiliary *haben* and a determiner *den*. They could simply consider grammatical a fragment of an utterance '*...hat den Ball... - ...has the ball...*', which might be a frequent combination in comparison with the fragment '*...kann den Ball... – ...can the ball...*'. Whereas the first fragment can occur often in short sentences directed to a child (e.g., *Das Kind* *hat den Ball – The child has the ball*), the last fragment can only be grammatical in larger and more complex syntactic constructions with an infinitive following the compliment (e.g., *Das Kind kann den Ball holen – The child can fetch the ball*). It is possible that infants in this experiment simply discriminated between these two fragments without paying attention to the use of the past participle in the end of the present perfect construction. Besides, infants could also rely on subtle prosodic cues. Thus, in the studies by Höhle et al. (2006) and by Santelmann and Jusczyk (1998), the non-adjacent dependencies in grammatical utterances were observed within phrasal units. There is a possibility that prosodic cues of phrasal boundaries assisted infants in tracking the non-adjacent dependencies.

A study by Van Heugten & Shi (2010) controlled for familiar sequences and prosodic factors. Here, French-learning infants were tested with passages of sentences containing grammatical and ungrammatical non-adjacent dependencies between French determiners and auxiliary verbs. The intervening element within a non-adjacent frame was represented by one consisting of bisyllabic nonce nouns. Such a pattern can be described as the following scheme: [[Determiner+ Noun]Auxiliary]. In grammatical trials, determiners and auxiliary verbs agreed in number: singular determiners preceded auxiliary verbs in a singular form (e.g., *La coupile va bientôt conduire – The_{sg} X $will_{sg}$ soon drive*). In ungrammatical trials plural determiners were incorrectly followed by singular auxiliaries (e.g., *Les coupiles va bientôt conduire – The_{pl} X $will_{sg}$ soon drive*). Furthermore, unlike in the studies by Santelmann & Jusczyk (1998) and Höhle et al. (2006), the non-adjacent elements span across a major phrasal boundary between the subject noun phrase and the verb phrase. Thus, the prosodic cues for phrasal boundaries did not align with the edge of the frame of the non-adjacent dependencies. Seventeen-month-old infants discriminated grammatical non-adjacent dependencies from ungrammatical ones, showing a looking preference for grammatical trials. These results demonstrate infants' knowledge of non-adjacent dependencies in their native language. Their processing of non-adjacent elements was not disrupted by intervening prosodic breaks. Infants showed generalized knowledge of non-adjacent grammatical relations, since the intervening words had never been heard before.

The variability of the intervening element can affect the tracking of a non-adjacent frame (Gómez, 2002). Gómez showed that low variability allows infants to pay attention to adjacent relations, and high variability enables them to track non-adjacent relations around the intervening element. She trained adults and eighteen-month-old infants with three-word AXC utterances from one of two artificial languages. The languages had a similar structure. The first and the third elements formed a non-adjacent relation (e.g., *pel-X-jic, dak-X-tood*). Such non-adjacent dependencies were unique to each of the languages. More specifically, in stimuli used with adults, *pel-X-rud, vot-X-jic* and *dak-X-tood* were non-adjacent dependencies in Language 1, whereas *pel-X-jic, vot-X-tood* and *dak-X-rud* were dependencies in Language 2. In stimuli used with infants, *pel-X-rud* and *vot-X-jic* were non-adjacent dependencies in Language 1, whereas *pel-X-jic* and *vot-X-rud* were dependencies in Language 2. During the training, participants heard utterances from only one of the languages. Across the experimental conditions, the variability of the middle element in the training varied. For adults, the variability ranged from 24 to 12, 6 and 2 X-elements. For infants, the training contained 24, 12, or 3 X-elements. Adults were exposed to the input with three different dependencies, whereas infants were trained with two dependencies. Test trials presented exemplars from both languages. They consisted of exactly the same artificial words as the training stimuli and were in exactly the same positions within the utterances (initial, middle, or final). Test stimuli from the same language as the training were grammatical; those from another language were ungrammatical. Since the middle elements were identical in both types of test stimuli, the difference was the dependency relation between the first and the last words of the triplet. Grammatical test stimuli had the same combination of the initial and the final elements as in the training, whereas in ungrammatical stimuli, the initial and the final items were combined according to the co-dependent relations for the other, untrained language. Thus, for participants who were trained with Language 1 dependencies, *pel-X-rud* and *vot-X-jic* utterances were grammatical, and *pel-X-jic* and *vot-X-rud* utterances from Language 2 were ungrammatical. The reverse was the case for participants trained with Language 2.

In Gómez (2002), both adults and infants discriminated between grammatical and ungrammatical utterances after having been trained with the language which had the highest variability of X elements (24 X elements). In the conditions with medium and low variability of X elements, participants did not discriminate between the trained and untrained languages. Gómez (2002) suggested that low variability of X elements in the training input guided participants' attention to adjacent dependencies between the first and second, and between the second and third elements. The initial and final bigrams of the test utterances (i.e., AX_ and _XB) had identical adjacent dependencies in both languages. Based on these identical adjacent relations, participants could not discriminate the two languages. When the variability of the middle element was the highest in the training input, the participants paid attention to non-adjacent dependencies. Gómez (2002) proposed that these differences in performance were linked to different values of transitional probability between adjacent items in the training input. When the transitional probability between adjacent elements in the input was high, learners were focused on adjacent dependencies (in the low and middle variability conditions). They did not go beyond learning those specific combinations and did not notice the non-adjacent items. When the transitional probability between adjacent elements was low, they tracked the non-adjacent dependencies (in the highest variability condition) rather than focusing on the specific bigrams of adjacent items.

A number of studies showed infants' early capacities to track non-adjacent dependencies in natural languages (e.g., Santelmann and Jusczyk, 1998; Höhle et al., 2006; Van Heugten & Shi, 2010). Adults were also able to track non-adjacent dependencies in the artificial language material (e.g., Newport and Aslin, 2004; Peña et al., 2002). Most of the studies reviewed above were focused on the tracking of specific non-adjacent elements. A generalization to novel items was examined mostly for the novel intervening middle elements, whereas non-adjacent frames were specific combinations learned in the training phase in laboratory settings or in the process of natural language acquisition. It is not entirely clear whether these non-adjacent frames are learned as specific item combinations or as abstract categories. Although the results are compatible with abstract category dependencies, there is no clear evidence supporting this interpretation.

In Marcus et al. (1999) and Gerken (2006), infants generalized identity-based patterns, including a non-adjacent dependency ABA pattern, to entirely novel stimuli. These studies suggest that infants were able to build abstract word categories after a brief exposure to artificial language input. However, linguistic categories in natural languages are more complex than identity-based categories in these studies.

A number of studies examined a syntactic categorization based on the non-adjacent dependencies in natural languages. Thus, in a study by Mintz (2002), adults considered a nonce word to belong to a category of other nonce words based on their common occurrence in shared non-adjacent dependency frames. A study by Saffran (2001) showed the learning of hierarchical syntactic category relations in an artificial grammar. However, due to the complexity of the grammar, it is not clear whether there is a specific contribution of non-adjacent dependencies in participants' learning.

1.1.4 Infants' generalization when two rules are possible

Sometimes input allows more than one interpretation of a rule. The study by Gerken (2006) examined nine-month-olds' interpretation of the input under the conditions when at least two generalizations were possible. In Experiment 1, as in the study by Marcus et al. (1999), infants were familiarized with a simple identity pattern (AAB for one group of infants and ABA for another group of infants). Two subsets of training stimuli were drawn from those used by Marcus et al. (1999). One subset of stimuli was used in the first condition where a syllable *di* was the B category word (here, *le le di*, *wi wi di*, *ji ji di*, *de de di* represented the AAB pattern, whereas *le di le*, *wi di wi*, *ji di ji*, *de di de* represented the ABA pattern). Another subset of stimuli was used in the second condition of Experiment 1 and had highly variable B category words (here, *le le di*, *wi wi je*, *ji ji li*, *de de we* represented the AAB pattern, and *le di le*, *wi je wi*, *ji li ji*, *de we de* the ABA pattern). Infants in both conditions of Experiment 1 were tested with novel items arranged in AAB and ABA sequences (i.e., *ko ko ba* and *po po ga* for the AAB test trials and *ba ko ba* and *ga po ga* for the ABA test trials).

Infants in the first condition (restrained B category) did not show any discrimination between the two types of test stimuli. Infants in the second condition (variable B category) showed a preference for the trained pattern, suggesting discrimination between the trained and untrained rules. The positive results in the second condition replicated those of Marcus et al. (1999) with seven-month-olds. The null results in the first, restrained B category, condition were of particular interest, since the training stimuli contained two patterns: one reflecting the position of identity items, and another based on the presence of the final *di* syllable. The null results could have two interpretations: the absence of any generalization, or infants' resistance to generalizing the learned pattern to test stimuli that did not use *di* as the B category. In a subsequent experiment, Gerken (2006) examined this latter interpretation.

In Experiment 2, infants were trained with the same familiarization stimuli as in the first condition of Experiment 1 (i.e., *le le di, wi wi di, ji ji di, de de di* for the AAB pattern, and *le di le, wi di wi, ji di ji, de di de* for the ABA pattern). Unlike in Experiment 1, test stimuli here contained the *di* syllable as the B word (i.e., *ko ko di* and *po po di* were used as AAB test trials, and *po di po* and *ko di ko* were used as ABA test trials). In Experiment 2, infants showed a preference for the trained pattern. These results suggested that they discriminated between the trained and untrained patterns, and that their generalized rule was abstract for A but specific for B (the *di* syllable). Infants applied the rule to novel sequences where words in the A position were novel whereas the B position was filled with the *di* syllable.

The training in the 1st condition of Experiment 1 (the restrained B category condition) and in Experiment 2 was identical. The positive results observed in Experiment 2 suggested that the training input was sufficient for infants to learn the rule. However, results in the two experiments were not the same: in Experiment 2, infants showed a discrimination of two test rules in the test, whereas in the 1st condition of Experiment 1, such discrimination was not observed. A crucial difference between these two conditions was in the test stimuli. In the 1st condition of Experiment 1, test stimuli were all novel items, whereas in Experiment 2, only A category words were novel, while the B category word was the same *di* as in the training input. Infants readily applied the rule to the test items sharing the same *di* words as in their training (in Experiment 2), but they did not apply the rule to entirely novel items (in the 1st condition of Experiment 1). The results suggest that infants learned a specific pattern: the rule (A-A-*di* for one group of infants and A-*di*-A for another group of infants) applies to items only if they share the same *di* word.

Interestingly, the training input in these two experimental conditions allowed two generalizations. The broader one could be generalized to any words. The more conservative one could be applied to new sequences only if the B word was the same one as in the training. Infants chose the conservative interpretation when the input lacked variability.

The sum of experimental results obtained by Gerken (2006) opens a possibility that input variability affects infants' interpretation of the input and generalization to novel instances. Two types of training were used in those experiments – the restrained B category training and the variable B category training. In the first case, B had zero variability, and in the second case, it had 100% variability. Infants showed a conservative interpretation of the restrained B category input and a broader interpretation of the variable B category input. The variability of the B category can be also presented by type-token variability. The restrained B category had one word type, and the variable B category had four types. The token frequency per type in the restrained set was 4 times greater than in the variable set. It is possible that it was this type-token variability, rather than merely the type variability of the B category, that guided infants' interpretations. However, type-token variability was not explicitly tested.

In a subsequent study by Gerken (2010), nine-month-olds were trained with the same stimuli that led infants to a narrower generalization in Gerken (2006), that is, the training stimuli with the restrained B category (i.e., *le le di, wi wi di, ji ji di, de de di* for the AAB pattern, and *le di le, wi di wi, ji di ji, de di de* for the ABA pattern). These triplets were presented repeatedly and randomly for two minutes. The AAB training set was presented to one group of infants, and the ABA training set was presented to another group of infants. Three counterexamples with the unrestrained B category were mixed among the last five stimuli of the training set. That is, *wi wi je, de de we, ji ji li* were mixed into the end of the AAB set, and *wi je wi, de we de, ji li ji* were mixed intro the end of the ABA set. After the training, infants were tested with novel triplets conforming to the trained and untrained patterns (*ko ko ba* and *po po ga* in AAB trials, and *ba ko ba* and *ga po ga* in ABA trials). The B category in the test stimuli was unrestrained. Infants discriminated between the trained and untrained patterns in the test. These results show that even few counterexamples in the training were sufficient for infants to make a broader generalization.

The study by Gerken (2006) confirmed the finding of Marcus et al. (1999) that very young infants can learn identity-based patterns and apply them to novel items. However, when the input allowed more than one generalization (more abstract versus conservative), they made the more conservative interpretation if the input contained low variability. Infants resisted the application of the learned rule to novel sequences that did not respect the properties of the input. When the training input contained high variability, they made a more abstract generalization. Such learning situations can be compared to the acquisition of natural languages. In natural languages, grammatical categories (e.g., nouns and verbs) contain a large number of word types. It is possible that type variability is necessary for abstraction. A lack of variability, on the contrary, should be beneficial for the learning of specific items, while impeding the abstraction.

1.2 Generalization, regularization and over-generalization

1.2.1 Defining the terms of generalization, regularization and over-generalization

This section will discuss three notions in the research on linguistic rule learning: generalization, regularization and over-generalization. Generalization is the application of the learned rule to novel items. For example, an English learner needs to extend his knowledge of the past tense ending *–ed* to other verbs which he has never heard before with that specific ending. Studies on generalization examine how learners extend the acquired rule to novel vocabulary or to new combinations of familiar vocabulary items.

The other two terms, regularization and over-generalization, imply the capacity to generalize, since they also concern the application of the learned rule to novel items. However, the focus in these two terms is shifted to the over-application of the rule. In other words, they concern cases where the application of the rule exceeds the properties of the examples. In the case of regularization, the rule is applied excessively to cases which are not overtly labeled as exceptions. So, the over-application is not necessarily an error, from the researcher's perspective. The term 'regularization' is often used in the context of creolization, i.e., the formation of consistent creole languages on the basis of inconsistent pidgin languages. It is also used in the context of the learning situation of a child who is exposed to imperfect and inconsistent grammar from adults, which he tends to regularize. In both contexts, a learner uses a rule more consistently than it is present in the input. An inconsistent input typically has more than one interpretation with regard to the application of the rule. For example, if a learner is exposed to a language where a determiner is only used with some nouns and not with others, he can make two interpretations: 1) only some novel nouns require the presence of the determiner, or 2) all novel nouns have to be used with that determiner. The second interpretation is what is called regularization: the learner applies the rule to all the cases, even those which might potentially not require the determiner. Since those cases were not explicitly heard with or without the determiner in the training, the researcher cannot consider such interpretation an overt error.

Over-generalization also means the over-application of the rule. In this case, however, the excessive use of the rule is considered an error. The term "over-generalization" is usually encountered in the field of first language acquisition when a child applies the rule wrongly to words which are exceptions in the adult language. For example, *feeled* instead of *felt* would be such an over-generalization error on the part of an English learning child.

In summary, all three terms – generalization, regularization and over-generalization – refer to the application of the rule to novel items. When a learner is exposed to entirely consistent input (i.e., the input with rule-conforming instances only), the application of the rule to novel items can be treated as simple generalization. Since the input is perfectly consistent, the over-generalization does not occur. When a learner encounters inconsistent input with some noise instances (i.e., utterances not conforming to the rule), the application of the rule to novel items can be treated as one of the three terms, depending on the researcher's focus of attention.

If the focus of attention is on the application of the rule, the term generalization is appropriate. It describes whether the learner is able to extend the rule to novel items in the presence or absence of the noise. If the focus, however, is on how the learner treats the noise instances, -- i.e., whether he treats them as exceptions or as ruleful -- then one of the other two terms should be applied (regularization or over-generalization). In the case of regularization, the 'noise' instances are often not defined as exceptions, and the learner applies the rule to the noise instances. Such application cases are not considered errors. As for over-generalization, it focuses on the overtly erroneous application of the rule to the 'noise' instances. These are exceptions to the language norm.

1.2.2 Generalization studies

Research on the generalization of rules to novel instances has already been reviewed extensively in section 1.1. Generalization studies are usually conducted using training, during which adults or children are exposed to an artificial language with an unknown grammar (e.g., Saffran, 2001; Wonnacott, Newport & Tanenhaus, 2008). Participants are subsequently tested with grammatical and ungrammatical utterances which contain either completely novel vocabulary, or at least novel combinations of the trained vocabulary. Several studies also have also reported generalization in preverbal infants with an artificial language (Marcus et al., 1999; Gerken, 2006) or with an unknown natural language (Gerken, Wilson and Lewis, 2005). Gómez and LaKusta (2004) demonstrated that infants could learn -- and generalize to novel instances -- adjacent relations between two word classes analogous to functors and content words in natural languages. Researchers have also examined infants' generalization in their native language (Höhle et al., 2004; Shi & Melançon, 2010; Cyr & Shi, 2013).

In natural language environments, infants are not exposed to perfectly systematic input. Gómez and LaKusta (2004) examined the extent to which infants could tolerate noise in their input and still learn and generalize the rule. By noise they meant the utterances in the learning input which violated the rule. Across experimental conditions, they manipulated the ratio of the rule-conforming instances and noise. One-year-olds were found to be able to tolerate up to 17% noise in their learning input; when noise reached 33%, their rule generalization failed. It has to be noted that the noise in this study was made with utterances violating the rule explicitly. Gómez and LaKusta (2004) manipulated the ratio of the rule and noise sentences by reducing the number of the rule-conforming sentences in the input. Thus, the perfect 100% training set contained 144 rule utterances. In the 83% training set, there were 120 rule-conforming utterances. In the 67% training set, there were only 96 rule-conforming utterances. Hence, apart from the rule and noise distribution, another factor that varied across experimental conditions was the overall exposure to rule-conforming sentences. An alternative to the design by Gómez and LaKusta (2004) would be the manipulation of rule and noise frequencies by adding noise to an unchanged number of rule sentences.

Another factor that could influence infants' learning in the study by Gómez and LaKusta (2004) is the type frequency. The 83% training set contained 20 types of rules and 4 types of noise. The 67% training set contained 16 types of rules and 8 types of noise. Rule-conforming instances were more frequent both overall and in type. Both frequencies could have an impact on infants' generalization.

1.2.3 Regularization studies

Most regularization studies emerged in an attempt to simulate creolization in the lab setting. Creolization is the formation of creole languages on the basis of pidgins. Pidgins are proto-languages used by speakers of two or more different languages. They appear in the situations when native speakers of different languages need to communicate daily and extensively but do not know one another's native language. Pidgins are often have inconsistent and omitted grammatical elements. Such omissions and inconsistency occur because of the conflict between two or more different grammars, as well as speakers' needs for simplicity in communication. Interestingly, pidgins do not stay incomplete and unstable. They evolve into creoles, which are languages with a complex and consistent grammatical structure. This evolution from the inconsistent and incomplete pidgin proto-language into the consistent and complex creole language demonstrates learners' linguistic capacities to regularize the inconsistent input they have heard.

We will focus on studies that examined what children and adults learn after being exposed to inconsistent input. A few studies examined regularization when participants are trained with artificial miniature languages in the lab. The training contained utterances in which the rule of interest was present with different degrees of consistency. Learners were further tested on their knowledge of the artificial grammar, with familiar and/or novel vocabulary. Hudson Kam and Newport (1999, 2005) trained adult participants with a new (Neg) VSO word order where nouns were followed by determiners. There were two determiners and two noun classes. Each determiner was assigned to a noun class. In Hudson Kam and Newport (2005), the determiners were assigned to noun classes in two different ways: on the arbitrary, 'gender', basis, and according to the count/mass categories of nouns (in these authors' 1999 paper, only one of these experimental conditions is reported). In addition, experimental conditions differed in the degree of consistency with which determiners were used in the input: they followed nouns from 45% of the time in the low consistency condition to 60% in the medium consistency condition, 75% in the high consistency condition, and 100% in the perfect consistency condition. Thus, there were eight conditions in total, differing by the consistency of determiner use in the input and by the noun classes assigned to the determiners. In the test, participants were required to carry out sentence completion and grammaticality judgments with trained vocabulary. Although the combinations of verbs and nouns in the test were novel, the combinations of nouns and determiners were heard during the training for some nouns which were used with determiners. Generalization was thus assessed from the learners' use and judgment for those combinations of familiar vocabulary. To examine whether the participants had a tendency to regularize the inconsistent examples, their use of determiners in the test was compared with the degree of determiner-consistency in the training set. No evidence of regularization was found:

adults' use of determiners in the test was no more consistent than had been in their training input.

Another study, however, shows that under some conditions adults do regularize inconsistent input (Wonnacott and Newport, 2005). This happens when an inconsistent rule is applied to completely novel items. Adult participants were trained with a miniature artificial language where 66% of sentences had a dominant VOS word order and 33% of sentences had a VSO order. Some nouns (labeled 'novel') were actually used in the vocabulary training, but only presented during the training within complete VSO sentences (in short two-word intransitive sentences). In the test, participants were asked to describe video scenes with the artificial language that they learned in the training. One group of participants produced sentences with the familiar vocabulary in exactly the same combinations heard during the training (Old Words group); another group produced sentences with 'novel' nouns and a completely new verb (New Words group); the third produced sentences of both types (Mixed group).

The analysis of the use of the dominant VOS word order did not generate much difference between the Old Words and New Words groups. Old Words participants reproduced the statistics of the input, using the dominant VOS word order with almost the same frequency as in the training set. Participants in the New Words group showed a slight degradation of the statistics of the training set, but did not differ strongly from the Old Words group.

However, when the results were re-analyzed according to the percentage of participants who used exclusively one of the word orders in all of their productions, 75% of the participants in the New Words group committed to one particular word order, with a larger number of participants selecting the dominant VOS order. In the Mixed group, an even larger number of participants showed regularization: 88% of participants consistently used the same word order with novel vocabulary, versus 38% of participants with the old vocabulary. By contrast, in the Old Words group, only 13% of participants committed to one word order in all of their productions. This substantial difference between the groups suggests that after being trained with inconsistent input, adults were more inclined to regularize novel vocabulary.

This pattern of results was also replicated in the second experiment of Wonnacott and Newport (2005) with an inconsistent use of a determiner. A structure of this artificial language was similar to the grammar used by Hudson Kam and Newport (2005). In this language, the training set had variable use of the *ka* determiner, which followed nouns 66% of the time, whereas the sentence structure was fixed as the VSO word order. In other respects, this experiment repeated the methodology used in the first experiment of Wonnacott and Newport (2005). The only difference was that the nouns used in the test phase had been presented during vocabulary training as separate words. Hence, those which were novel were never encountered with the determiner during the training. Here, three types of productions were considered as regularization of the variable input: using or skipping the determiner consistently after all subjects, after all objects, or after all subjects and objects. Again, 77% of participants in the New Words group showed one of these three patterns, whereas only 23% of participants in the Old Words group committed to one of the patterns. The Mixed group, again, showed stronger generalization even with the familiar vocabulary: 69% were regularizing with familiar vocabulary, and 77% were regularizing with novel vocabulary.

Overall, the study by Wonnacott and Newport (2005) suggests that adults regularize more when applying an abstract rule to novel items. Even the regularization to familiar items was strengthened when participants had more experience with novel vocabulary (as in the Mixed group).

The study shows a similar pattern of adult performance with two different rules: a word order and a determiner use. In the first case, the inconsistency in the input was introduced through two different word orders. In the second case, the inconsistency was in the use and omission of a determiner. In both cases, whether choosing between two different uses, or between use and non-use, adults tended to regularize with novel vocabulary. These findings are consistent with the study of Hudson Kam & Newport (2005), where adults did not regularize inconsistent input to familiar vocabulary when the inconsistency was based on the use and non-use of a determiner.

Wonnacott and Newport (2005) joined different kinds of regular uses together: a systematic application of one of two rules in the input (in the case with two word orders), and a systematic use or non-use of the grammatical element (in the case of the determiner use). In the experiment with two word orders, not all systematic users in the New Words group adopted the word order that had been dominant in the training set. A third of them systematically used the non-dominant word order. It was not clear what produced these individual differences, and whether both of them represented the same kind of regularization. Unfortunately, no information is available about the percentage of systematic users who adopted the dominant and non-dominant word order in the Mixed group of participants. Similarly, there is no description of how many participants in the experiment with the determiner use exhibited each of the possible regularization behaviors: a systematic use of the determiner with all nouns, or all subjects, or all objects; a systematic non-use of the determiner with all nouns, or all subjects, or all objects. Since all these behaviors were united under the unique label of regularization, it is not possible to examine whether the regularization included a lot of cases of the non-dominant pattern in the training set (i.e., the non-use of determiners).

There exists evidence that children can regularize beyond the statistical properties of the training input, and do so for both novel and familiar vocabulary

(Hudson Kam & Newport, 2005). Recall that adults did not regularize when tested with the familiar vocabulary. However, 5- and 7-year-olds did show regularization in the experiment. Together with a new group of adults, children were exposed to a miniature artificial language with a (Neg) VSO word order. In that language, nouns were followed by determiners. The language was simplified by having only one determiner and one noun class to precede that determiner. Researchers introduced two experimental conditions: the condition with perfectly consistent training where the determiner was used 100% of the time, and the inconsistent input, where the determiner was used only 60% of the time. Children performed sentence completion and grammaticality judgment tasks with familiar vocabulary.

In the sentence completion task, children and adults showed different patterns in reproducing the use of determiners in the training. Adults, overall, repeated the same pattern as the one found in the first experiment of this study, although here (in Experiment 2) the language was less complex and the number of participants was smaller. In the consistent condition, all the adults adopted a systematic use of the determiner in their productions. In the inconsistent condition, only 50% of adults showed consistent linguistic behavior; that consistent behavior was the systematic omission of the determiner.

Among children, however, in both consistent and inconsistent conditions, over 70% demonstrated perfect consistency in their productions. In the consistent condition, most of those children were systematic users of the determiner, whereas in the inconsistent condition, the majority of children with a consistent linguistic pattern were systematic non-users of the determiner.

These results suggest that children can produce patterns more systematically than those present in the input, despite their good knowledge of consistent determiners, as shown by grammaticality judgment tasks. This is the first study in a laboratory setting that suggests that children can regularize the inconsistent training input. It is, however, unclear, why some children with consistent linguistic behavior chose a pattern opposite to the more frequent one in the training input. Thus, 25% of children in the perfectly consistent condition chose not to use the determiner at all, and did so systematically. The percentage of participants who systematically omitted the determiner in the inconsistent condition reached 57%, despite the fact that the determiner omission was not a dominant pattern in their input: 60% of utterances of the input did contain the determiner. Such a high number of systematic non-users raises the question of whether their knowledge of the determiner use was as robust as the knowledge of systematic users. Hudson Kam and Newport (2005) mention that the results they obtained differed from their expectations: they anticipated that regularization, if any, would manifest itself through a more systematic use of the determiner. The percentage of systematic users among children, however, only reached 50% in the consistent condition, despite the fact they were exposed to the perfectly consistent input where determiners were used all the time. This was markedly different from adults: all adults reproduced the perfectly consistent input in all their productions. The percentage of systematic users among children in the inconsistent condition was even lower than in the consistent condition, as low as 14.3%. This reluctance on the part of children to use determiners systematically suggests that they did not regularize the variable input. It is known that in the early stages of first language acquisition, children often omit determiners in their productions, despite having knowledge of determiners, which they demonstrate in perception studies. It is possible that the systematic non-users in the study by Hudson Kam and Newport (2005) were simply exhibiting the common behavior of dropping the determiner even from the perfectly consistent input. The dropping was less pronounced in the perfectly consistent condition, where only 25% of children omitted the determiner systematically, and it became a dominant trend in the inconsistent condition. In the inconsistent condition, the input already contained 40% of utterances without the determiner, and 57% of children chose to omit the determiner systematically. It is possible that the systematic non-use cannot be considered pure regularization, given children's common tendency not to use determiners in their productions. This question, however, stays open.

The expectation by Hudson Kam and Newport (2005) that in cases of regularization, if any, children would use the optional determiner more systematically was based on research conducted with a rare participant, named Simon (Ross & Newport, 1996; Newport, 1999; Singleton & Newport, 2004). Simon was a congenitally deaf child who acquired his first language, ASL, from the imperfect and inconsistent signing of his deaf parents who learned ASL in their late teens. Simon's productions appeared to be more systematic than his parents' inconsistent input. That process of changing the inconsistent input into the systematic grammar was first called "frequency boosting" and "regularizing" (Singleton & Newport, 2004).

Both Simon's and his parents' productions were recorded in spontaneous interactions and elicited productions over a seven year period beginning when Simon was 2 years old. Simon's elicited productions at the age of 7 were directly compared to those of his parents and to a group of deaf children whose parents were native deaf signers (Singleton & Newport, 2004). The focus of the researchers was on his acquisition of morphology in motion verbs. The morphology of motion verbs in ASL is very complex and difficult to learn for non-native signers. Simon's parents hence demonstrated an imperfect use of morphemes in motion verbs. However, Simon's performance was comparable to children who learned ASL from perfectly consistent input on most of the morphemes; for some morphemes, Simon's performance was even better. All those morphemes which received the "frequency boosting" were present in his parents' input with 65% to 76% consistency. Simon used them in 84 to 91% of his productions. For one morpheme, Simon did not reach the performance of other children, but still showed an improvement over his parents' 37-43%, achieving 59%. Only in the use of one morpheme did his performance stay at his parents' level of consistency. The authors suggested that the degree of inconsistency in parents' input affected Simon's regularization, so that the frequency boosting was more modest or even absent when the inconsistencies in the input were stronger. However, Singleton and Newport (2004) also suggested that Simon's difficulty with some morphemes could be explained by other factors, related to the difficulty of the acquisition of those particular morphemes even by children who learn ASL from native signers.

1.2.4 Over-generalization errors in children's productions

This section will briefly outline evidence that children make over-generalization errors in their productions during language acquisition. A review of over-generalization errors in children's productions can be found in Bowerman (1988) who summarized her observations of her own two English-speaking children. They made over-generalization errors in various grammatical constructions.

There is evidence of over-generalization errors in dative alternation. According to the rules of the English language, some verbs can be equally used in a prepositional indirect-object construction (e.g., *The girl gave a pen to the teacher*) and in a double-object construction (e.g., *The girl gave the teacher a pen*). Other verbs do not allow such alternation between two constructions. For such verbs, only the prepositional indirect-object construction is grammatical. Bowerman (1988) gives such examples of children's over-generalization errors. A 3;1-year-old produced a sentence '*I **said her** no*'. Here, the double-object construction is used incorrectly with the verb '*say*' which only allows a prepositional indirect-object construction. A correct sentence would be '*I said no to her*'. A two-and-a-half-year-old produced a sentence '*Don't **say me that** or you'll make me cry*'. Again, the verb '*say*' was used incorrectly in the double-object construction. A grammatical sentence would be '*Don't say that to me or you'll make me cry*'. Over-generalization errors were also observed with the verb '*choose'*, in a production of a two-and-a-half-year-old (*'I want Daddy **choose me** what to have'* instead of the grammatical *'I want Daddy choose for me what to have'*) and in a production of a 5-year-old ('***Choose me the ones** that I can have'* instead of the grammatical *'Choose for me the ones that I can have'*). Another example is a production of a 7;8-year-old: '*Shall I **whisper you something**?*' instead of the grammatical '*Shall I whisper something to you?*'.

Bowerman (1988) also reports the observations of over-generalization errors in passive constructions. Among these errors is the incorrect use by a 3;8-year-old of an irregular verb in the past participle formed as if it were a regular verb: '*Both are going to be **go-ened** in'* instead of the grammatical 'Both are going to be gone in'. Other errors involve using incorrectly causative verbs that should be used in the active voice constructions. For example, at 3;6 a child produced a sentence '*Until I'm four I don't have to be **gone***', instead of the grammatical '*Until I'm four I don't have to be taken to the dentist*'; at 4;3, a child produced a sentence '*Why is the laundry place **stayed** open all night?*' instead of the grammatical '*Why is the laundry place kept open all night*?'. A 4-year-old made a sentence '*He's gonna **die** you, David. The tiger will come and eat David and then he will be **died** and I won't have a brother any more*', instead of the correct '*...then he will be killed...*'. Other errors involve forming novel past participles out of nouns and adjectives. Thus, a three-and-a-half-year-old, referring to crackers in a bread box, produced a sentence '*If you don't put them in for a very long time they won't get **staled***', instead of the grammatical '*...they will get stale*'. At 5;2, a child made a similar error forming a past participle out of a noun: '*Mommy will get **lightninged***' instead of the grammatical '*Mommy will get struck by lightning*'.

Other over-generalization errors summarized by Bowerman (1988) involve causativity. For example, at 2;8+, a child said '*I don't want any more grapes; they just **cough** me*', instead of the grammatical '*...they make me cough*'. Between the ages of three and four, such errors were observed: '*Don't **giggle** me*' instead of '*Don't make me giggle*'; '*Will you **climb me up** there and hold me?*' instead of '*Will you help me climb up there and hold me?*'; '*I'm gonna put the washrag in and **disappear** something under the washrag*' instead of '*...and make something disappear under the washrag*'; '*Did she **bleed** it?*' instead of '*Did it make her bleed?*'. At five and six, similar errors were observed: '*I want to **comfortable** you*' instead of '*I want to make you comfortable*'; '*Do you want to see us **disappear** our heads?*' instead of '*Do you want to see our heads disappear?*'.

More over-generalization errors were found in locative alternation constructions (Bowerman, 1988). Some verbs in English can be used in both locative alternation constructions (e.g., '*spread butter on the bread*' and '*spread the bread with butter*'). Other verbs, however, can only be used in the first or the second construction. For example, the verbs '*pour*', '*spill*', '*steal*' can only be used in the first construction, and the verbs '*fill*', '*cover*', '*rob*' can only be used in the second construction. Bowerman (1988) observed in her children the incorrect use of such verbs specific to one of the constructions. Thus, between the ages of four and five, children used the verbs of the second construction in the first construction: '*I'm gonna **cover** a screen over me*' instead of '*I'm gonna cover myself with a screen*'; '*Can I **fill** some salt into the bear?*' instead of '*Can I fill the bear with some salt?*'; '*She's gonna **pinch** it on my foot*' instead of '*She's gonna pinch my foot with it*'. Similarly, they also used the verbs of the first construction in the second construction: at 2;11, '*Mommy, I **poured** you. (...) Yeah, with water*' instead of '*Mommy, I **poured** the water on you*'; at 4;11, '*I don't want it because I **spilled** it of orange juice*' instead of '*I don't want it because I spilled orange juice on it*'.

Over-generalization errors reported by Bowerman (1988) also included reversative un-prefixation. Children were making novel words with the reversative un- prefix: '*How do you* ***unsqueeze*** *it?*' (3;11); '***Uncapture*** *me!*' (3;10); '*And I'm never going to* ***unhate*** *you or nothing!*' (4;7); '*And* ***unstraighting*** *it?*' (4;5); '*He tiptoed to the graveyard and* ***unburied*** *her*' (5;1); '*I'm gonna* ***unhang*** *it*' (7;11).

1.3 Models of exception learning

Natural languages usually contain exceptions to grammatical rules. Some cases are overt exceptions, i.e., their correct form is clearly different from the rule-governed cases. The overt exceptions present an alternative form or construction. The alternative form/construction can be encountered in adults' speech and can thus be learned even from positive evidence alone. For example, an extended exposure to the past form of irregular verbs – *ran, ate* – would be sufficient for a child to override a possible over-generalization error – **runned, *eated.* In addition, there are non-application exceptions for which a dominant rule does not apply. These do not have an alternative form/construction. Among such non-application exceptions are English verbs that do not allow the optional dative alternation movement. For example, although *tell* and *say* can both be encountered in identical constructions (*Dad told a story to Sue*; *Dad said something nice to Sue*), the verb *say* cannot go through the dative alternation (e.g., **Dad said Sue something nice* would be ungrammatical), whereas the dative alternation can apply to the verb *tell* (e.g., *Dad told Sue a story*). An important question is how children learn such exceptions and resist their over-generalization.

A particular challenge for learning of non-application exceptions is the well-known fact that children's grammatical errors are almost never corrected; and, even when they are, children do not tend to be influenced much by such corrections. In the case of overt exceptions, children can hear the correct alternative in adult speech and eventually notice that they are not using the adults' form. With non-application exceptions, infants hear only the correct form in adult speech. They receive no indication that the alternative they produce is ungrammatical, since there is no grammatical alternative in adults' speech. The problem of how children generalize the acquired grammar to novel instances without over-generalizing it to lexical exceptions is known as a Baker's paradox (Braine, 1971; Baker, 1979).

Braine (1971) suggested a 'Discovery Procedures' Acquisition Model to address a question of rule and exception learning. The model was based on two major components: a *scanner* and a *memory component*. The *scanner* component referred to the ability to notice some patterns in the linguistic input. Afterwards, these patterns were stored in the *memory component*. The *memory component* contained some *intermediate stores* and a *permanent store*. The frequency of such patterns was crucial for the model. Patterns encountered more frequently would move to the *permanent memory store*, via the *intermediate stores*. The less frequent patterns stayed in the intermediate stores. This model was hypothesized to be immune to unsystematic errors: the unsystematic errors could never reach the *permanent memory store* and, hence, could not affect the learning of linguistic patterns. According to the model, the more general patterns would be learned first, since they occur in a larger number of sentences. More specific patterns would be learned after the more general patterns. At that later stage of acquisition, children learn which syntactic contexts are more appropriate for specific items. At this stage, the knowledge of linguistic rules can be adjusted by exceptions.

Braine's model assumes that a child can make over-generalization errors at the early stages of learning, particularly in non-application exceptions. At this point, their permanent memory store contains general rules, for example, the dative alternation rule. By that rule, verbs can occur in two constructions with identical meaning (e.g., *Dad told a story to Sue - Dad told Sue a story*). Children can construct an over-inclusive grammar by using all verbs in both constructions, even if for some verbs one of these constructions would be ungrammatical (e.g., *Dad said something nice to Sue - *Dad said Sue something nice*). Later, the scanner encodes the properties of the co-occurrence of specific verbs with each construction. Specifically, it encodes the fact that the verb *tell* often occurs with both constructions, whereas the verb *say* occurs frequently in the first construction only. The differential frequency in the use of specific items allows a child to overcome his initial over-generalization errors. Therefore, in Braine's induction-based model, over-generalizations occur at the early stage and are later inhibited.

Baker (1979) also assumed that input does not contain negative evidence. He argued that non-application exceptions cannot be learned by induction from the input, since both positive and negative evidence about this kind of exception is absent. He therefore concluded that an induction based view cannot explain the learning of these exceptions. In the case of non-application dative alternation exceptions, Baker assumes that children do not commit over-generalizations. According to his Innate Constraint Model (Baker, 1979), only grammatical regularities which can be learned from positive evidence alone belong to syntactic structures, while exceptions to the rule should be treated as constraints from the lexicon. This model assumes that no general syntactic rules involving non-application exceptions need to be learned. Thus, the problem of exception learning does not exist according to this model. Children only produce the exception lexical items which they have heard before in the same syntactic context. Applying this model to dative alternation, Baker (1979) suggests the following steps. An English-learning child encounters a number of sentences containing two structures: *John gave the book to Alice. John gave Alice the book. We sent a letter to him. We sent him a letter. George said something uncharitable to Maxine. We reported the accident to the police.* According to Baker's model, a child does not derive one type of dative construction from the other with a transformational rule. Instead, he acquires two phrase structure rules. Afterwards, he attributes each individual verb to appropriate structures, relying on subcategorization features that indicate the environment in which the verb should appear. The attribution of individual verbs to one of the structures takes place after a child encounters that verb in that specific structure. The weakness of Baker's argument was in the assumption that children do not make over-generalization errors in such non-application exceptions as dative alternation.

Bowerman (1988) later showed that children do in fact make dative alternation over-generalizations. Production and grammaticality judgment studies in English have shown that children make over-generalization errors in dative alternation, causative verb formation, passivization, locative alternation and un-prefixation (Bowerman, 1974, 1982a, 1982b, 1983, 1988; Mazurkewich and White, 1984; Hochberg, 1986). Children also use novel verbs in passive, alternated datives and causative forms even though they have never encountered those verbs in such syntactic contexts (Pinker et al., 1987; Pinker, 1987; Maratsos et al., 1987).

According to the Criteria Approach model (Mazurkewich and White, 1984; Pinker, 1984; Pinker, 1987), the application of a rule to a word depends on whether it shares semantic, morphological and phonological criteria with other words. For example, Mazurkewich and White (1984) suggested that in the dative alternation rule, a combination of semantic and morphological properties of a verb define whether the verb could be used with double object compliments (e.g., *John gave Fred the book*). For some verbs, such use is grammatical, as is their use with prepositional phrase compliments (e.g., *John gave the book to Fred*). Such verbs are alternating. For other verbs, only the use of prepositional phrase compliments (e.g., *John reported the accident to the police*) is grammatical, whereas the use of double object compliments (e.g., **John reported the police the accident*) is ungrammatical. Mazurkewich and

White (1984) suggested that the distinction between alternating and non-alternating verbs depends on a combination of their semantic and morphological properties. Morphologically, most English alternating verbs are 'native' (that is, of Anglo-Saxon origin), whereas most non-alternating verbs are of Latin origin (Green, 1974; Oehrle, 1976; Stowell, 1981). Semantically, English dative alternative constructions contain an animate indirect object that is the 'prospective possessor' of the direct object (Goldsmith, 1980; Stowell, 1981). For example, '*Joe Smith*' is a 'prospective possessor' of '*five bucks*' in a grammatical sentence with dative alternation: *I owe Joe Smith five bucks*. In a sentence *I owe Joe Smith this example* the animate indirect object '*Joe Smith*' is not a 'prospective possessor' of '*this example*' since an example cannot be possessed. Mazurkewich and White (1984) suggested that children's over-generalization errors in dative alternation are caused by their inability to integrate both morphological and semantic factors. The relatively less common observations of dative alternation over-generalizations in children can be due to the fact that children know fewer Latinate words at an early age. Mazurkewich and White (1984) observed dative alternation over-generalization with older children. In their study, children made grammaticality judgment on sentences with dative alternation, particularly, sentences containing a preposition 'to' in a prepositional phrase compliment (e.g., *Bob reported the accident to the police*). Among those sentences were ungrammatical sentences with Latinate verbs (e.g., **Bob reported the police the accident*). The 9-year-olds judged 46.7% of such ungrammatical sentences to be grammatical. The 12-year-olds judged 33% of them to be grammatical, whereas for 16-year-olds only 11% of them seemed grammatical. It is possible that the results are related to the fact that older children might have heard more Latinate verbs, which do not alternate. Mazurkewich and White (1984) suggested that younger children initially over-generalize more, but subsequently retreat from over-generalization errors after discovering the criteria for exceptions. First, they discover the semantic requirement that an indirect object in dative alternation, which is a goal or beneficiary of the direct object, should also be a possessor of that direct object. Later, they constrain this knowledge by the morphological requirement that Latinate verbs do not alternate.

The Criteria Approach Model was criticized by Bowerman (1988). She questioned why children needed to make the effort to restrain their over-general grammar, since adult productions in one construction (rather than alternating) are compatible with this over-general grammar. She further argued that some cases do not work according to the criteria proposed by Mazurkewich and White. The cases she noted did belong to the alternating class in terms of semantic and morphological criteria, but did not allow a dative alternation (e.g., **I chose you a book at the library sale*).

Other authors have speculated about the role of preemption. According to the Preemption model (e.g., Clark, 1987; Markman, 1989; Pinker, 1984; Goldberg, 1995), the erroneous over-generalization construction used by children is in direct semantic competition with the correct form of an exception used by adults. For example, an English-learning child can make an error *go-ed* instead of *went*, mistakenly over-generalizing a regular past tense inflection to an irregular verb. However, in adult speech, the child hears the form *went* that refers to the same meaning. Since both *go-ed* and *went* can be used to communicate the same thing, the highly frequent adult form over-rides the children's error, preempting it. Studies with 4.5-year-olds and 6-year-olds have shown that children can use such frequency information in their language acquisition (Brooks & Tomasello, 1999; Brooks & Zizak, 2002). For example, Brooks & Tomasello (1999) taught 2- and 3-year-olds to produce nonce verbs in passive and active transitive constructions. Children were asked to answer questions in a play session. In their productions, children tended to generate constructions identical to those which were primed in their training. That is, a nonce verb that was introduced to children in the passive construction was later produced more often in passive sentences. The fact that children in Brooks &

Tomasello (1999) were 2- and 3-year-olds and not infants at the early stage of language acquisition suggests that this process requires a comparison of semantic properties and pragmatic contexts on top of the frequency of use. This can be a far too complex task for younger children.

The third, Entrenchment, model is a learning account independent of the semantic and pragmatic aspects of the acquired input. The frequent use of a word in a certain construction "entrenches" its use to this construction, so that the word is less likely to become a subject of an over-generalization error for the wrong construction (Braine & Brooks, 1995). Brooks et al. (1999) observed this effect with children who were less likely to make over-generalizations with high-frequency English verbs than with low frequency English verbs. In a game situation, 3-, 5- and 8-year-olds heard utterances containing several transitive and intransitive verbs. Half the verbs were high frequency verbs, the other half were low frequency verbs. Afterwards, in a game situation, children were encouraged to produce utterances with those verbs. Children's over-generalization errors were more likely with low frequency verbs. In addition, children were more likely to use intransitive verbs in transitive constructions than to use transitive verbs in intransitive constructions. The effect of age was not significant, although the percentage of errors was slightly lower in older children.

In studies where children and adults were asked to give grammaticality ratings for over-generalization errors, they tolerated the over-generalization errors more for low frequency verbs (Theakston, 2004; Ambridge et al., 2008). Thus, Theakston tested 5-year-olds, 8-year-olds and adults with high and low frequency verbs used ungrammatically in argument structures (e.g., **I spilled the carpet with juice* – with a high frequency verb *spill*; **I dripped the table with milk* – with a low frequency verb *drip*). The participants were asked to judge whether they found the sentences to be grammatical, and to rate their degree of grammaticality. In all groups, the erroneous sentences with low frequency verbs were judged as more grammatical than the erroneous sentences with high frequency verbs. Ambridge et al. (2008) tested 5-6-year-olds, 9-10-year-olds and adults with a similar grammaticality judgment task. They asked the participants to judge the grammaticality of high and low frequency verbs used in grammatical intransitive constructions (e.g., *Bart fell into a hole* – with a high frequency verb; *Lisa tumbled into a hole* – with a low frequency verb) and ungrammatical transitive ones (e.g., **The man fell Homer into a hole* – with a high frequency verb; **The man tumbled Bart into a hole* – with a low frequency verb). All participants had a stronger preference for the grammatical use of high frequency verbs than for low frequency verbs.

The problem of 'embarrassing' exceptions can be extended to all words which are not yet encountered by a child in either regular or irregular form/construction. If the positive evidence has not had a chance to appear, how will a child know whether this word can be generalized loosely to a regular rule, or if it should rather be conservatively treated as a potential exception?

The role of frequency in exception learning was demonstrated by Wonnacott, Newport and Tanenhaus (2008). They trained adult participants with a miniature artificial language. Verbs were presented in two verb argument structures (e.g., four verbs occurred in one structure, and four verbs occurred in another structure). The frequency of the verbs was manipulated (e.g., among four verbs used in one structure, two verbs had high frequency, and two verbs had low frequency). In the test, participants produced high frequency verbs in a structure where the verbs occurred in the training. They made more errors with low frequency verbs. These results suggest that a high number of repetitions leads to entrenchment.

With young infants, the role of frequencies in the learning of exceptions was not tested directly. Indirect evidence comes from a finding that high token frequency leads to the learning of specific items (Gerken, 2006). Infants were trained with three-word utterances that contained a final word with a high frequency of occurrence. Infants learned to track that frequent final word. Other indirect evidence comes from a study on non-adjacent dependencies (Gómez, 2002). When a middle element occurred frequently, infants learned specific sequences.

1.4 Type and token frequencies in abstraction of rules and learning of specific examples

The roles of type and token frequencies in abstraction of rules and learning of specific examples were addressed in a study by Wonnacott, Newport and Tanenhaus (2008). In this study, adults were trained with a miniature artificial language containing two verb arguments structures: Verb Agent Patient structure (e.g., *Glim tombat blergen* meaning *The giraffe hit the li*on) and Verb Patient Agent Particle structure (e.g., *Glim blergen tombat ka* meaning *The giraffe hit the lion*). In one experimental condition labelled the *Lexical language* (further on we will call it the Lexicalist language), each verb was assigned to one of the structures. In particular, seven verbs were used in one structure, and one verb in another structure. Hence, the first structure was seven times more frequent in type and in overall frequency. When tested with novel verbs, participants used them more in that frequent structure. Similar results were found with children with another kind of a Lexicalist artificial language (Wonnacott, 2011). English-speaking six-year-olds were trained with artificial language sentences containing an English noun that was followed by a nonce particle (*tay* or *dow*). They saw pictures of two identical cartoon animals, heard sentences and repeated them aloud. For example, children saw a picture of two pigs and heard a sentence *moop pig tay*. Given the context, children could implicitly learn that *moop* meant *there-are-two* and the sentence meant *There are two pigs*. The frequency of particles was manipulated, while the structure of sentences was the same (i.e., Non-word Noun(English) particle). In one experimental condition labeled the Lexicalist condition, three nouns were used with only one particle, and one noun was used with only another particle. Since each of the nouns occurred equally frequent in the training, one particle was used more frequently than another. That is, utterances containing nouns with one particle (e.g., *tay*) occurred more often than utterances with another particle (e.g., *dow*). In the test with novel nouns, children were shown pictures with animals that were not used in the training. They were encouraged to produce artificial language sentences with English nouns corresponding to those animals. They produced sentences starting with *moop* more with the particle that was more frequent in the training. For example, when they were trained more with *tay* utterances, they produced new sentences with novel nouns mostly with *tay* particle. Thus, they made a generalization based on a more frequent particle. One possible interpretation of results by Wonnacott, Newport and Tanenhaus (2008) and Wonnacott (2011) is that high type frequency supports generalization. However, in both studies the type frequency was not separated from overall frequency.

The same studies by Wonnacott, Newport and Tanenhaus (2008) and Wonnacott (2011) also included a different type of input training, which they called the Generalist language. In the Generalist language, all target words occurred in both patterns. For example, Wonnacott, Newport and Tanenhaus (2008) presented to adults such sentences as *Glim tombat blergen* (VPA structure) and *Glim blergen tombat ka* (VAP_ka structure). And in Wonnacott (2011) chldren heard such sentences as *moop pig dow* (a sentence with one particle) and *moop pig tay* (a sentence with another particle). In both studies, one pattern was more frequent in the training. In the study with adults, each verb was presented in the VPA_ka structure seven times more often than in the VAP structure (Wonnacott, Newport and Tanenhaus (2008). In the study with children, each noun occurred three times more often with one particle than with another particle (Wonnacott, 2011). When tested with novel words, both children and adults used them in that dominant pattern more than in another pattern.

Do learners over-generalize the trained words from the minority pattern to the dominant pattern? The experiments by Wonnacott, Newport and Tanenhaus (2008) and Wonnacott (2011) give some clue to this question. In particular, there were additional minimal exposure items after the regular training. They were four additional words, each presented to participants four times. Two words occurred only in the more frequent pattern, and two other words occurred only in the other less frequent pattern. The last two minimal-exposure words, which were presented to participants only in the less frequent pattern, are of particular interest. In the Lexicalist condition, those minimal exposure words that were trained in the minority pattern were produced predominantly in the same minority pattern by both children and adults. Thus, they were not over-generalized to the other overall dominant pattern. Children made more over-generalizations but they still produced the minimal exposure words from the minority pattern predominantly in the same minority pattern where they were heard, i.e. with the same, less frequent particle as during their initial presentation (Wonnacott, 2011).

The studies by Wonnacott, Newport and Tanenhaus (2008) and Wonnacott (2011) did not show clear evidence of over-generalization to the dominant pattern of minimal exposure words presented to participants solely in the minority pattern. The reason that the learners did not over-generalize may be that the frequency of those words was sufficiently frequent. In a different experiment, Wonnacott, Newport and Tanenhaus (2008) specifically tested weather a higher number of occurrences of target verbs would lead to a better learning of their specific patterns. The training contained four verbs used in one structure and four verbs used in another structure. In each structure, two verbs had high frequency, and two had low frequency. Besides those, there were four alternating verbs, each occurring equally in both structures, and the frequency of those verbs was not manipulated. Overall, neither structure was more frequent. In the test, participants produced high frequency verbs in the structure where the verbs occurred in the training. They made more over-generalization errors with low frequency verbs. They still used them in the majority of productions correctly, i.e. in the same structure where they occurred in the training. However, the number of correct productions was significantly lower for low frequency verbs than for high frequency verbs. As for the alternating verbs, they were almost equally produced in both structures. Since their frequency was not manipulated, participants showed probability matching. They did not show entrenchment.

These results suggest that a high token frequency leads to learning of specific examples. The learning of specific examples is related to the question of over-generalization. According to the idea of entrenchment, the higher the token frequency of an exception, the more likely it is learned as a specific instance and resists over-generalization. The lower its token frequency, the more likely it will be over-generalized. Indeed, with high frequency verbs children make less over-generalization errors and tend not to accept others' over-generalizations (Brooks et al., 1999; Theakston, 2004; Ambridge et al., 2008). Over-generalization errors in this context are the misuse of a verb to a wrong structure.

Overall, the studies by Wonnacott, Newport and Tanenhaus (2008) and (Wonnacott, 2011) suggest that children and adults track both statistics of general patterns and of occurrence of specific examples in those patterns. Within a probability-based framework of Bayesian learning, a model was developed in order to simulate learning of an artificial miniature language in adults. In particular, a Hierarchical Bayesian model, developed by Kemp et al. (2007), was applied by Perfors, Tenenbaum & Wonnacott (2010) to the study of Wonnacott, Newport & Tanenhaus (2008). This model tracks both the statistics of the use of each particular example in one or another structure, and the statistics of the overall frequency of that structure. If the overall frequency of one structure is high, the model learns that a general distribution is a reliable cue for the use of a structure. And if specific examples occur consistently with one structure, then the model learns the use of those examples in that particular structure. The exact nature of frequency and consistency was not precisely defined.

The Hierarchical Bayesian model (Perfors, Tenenbaum & Wonnacott, 2010) reproduced the results of both Generalist and Lexicalist grammar in Wonnacott, Newport and Tanenhaus (2008). For example, it reproduced the Lexicalist learning condition where the high type and high overall frequency of one structure led participants to generalize novel verbs to that dominant structure. According to the results of Wonnacott, Newport & Tanenhaus (2008), specific examples are learned and hence avoid over-generalization when they occur frequently. Even four repetitions were sufficient for learners (Wonnacott, Newport and Tanenhaus, 2008). The simulations by Perfors, Tenenbaum & Wonnacott (2010) showed results consistent with human learning: when statistics were high for the number of repetitions of specific examples, they were learned in a specific structure and hence avoided being over-generalized. It is not clear what prediction the model would make when the type frequency of rule exemplars is low but each is repeated frequently to yield a high overall frequency. In Wonnacott, Newport & Tanenhaus (2008), type frequency was not separated from the overall frequency. Therefore, the simulations did not include the test of this factor.

1.5 Infants' sensitivity to morphological markings

In natural languages, the grammatical category of a word can be inferred from its inflectional morphemes. For example, in Russian, an ending *-aya* defines a word as an adjective referring to a noun of a feminine gender.

Studies on the comprehension of meaning in infants show early sensitivity to morphological markings. Thus, Waxman and Booth (2001) demonstrated that English-learning 14-month-olds treat pseudo-words as either nouns or adjectives depending on their grammatical form. When a pseudo-word occurred in the familiarization with a plural ending *–s* (e.g., *These are blickets*) and with an article *a* (e.g., *This one is **a** blicket and this one is **a** blicket*), infants interpreted it as a noun signifying an object category (e.g., animals). And when the pseudo-word contained an adjective suffix *-ish* (e.g., *These are blick**ish**. This one is blick**ish** and this one is blick**ish***), infants interpreted it as an adjective signifying a property of objects (e.g., purple objects). Oshima-Takane et al. (2011) showed that Japanese-learning 20-month-olds can learn to map a pseudo-word to an action. In familiarization and test sentences, the nonce word was followed by a Present Progressive of *do* (***shite(i)ru*** – *do**ing***). This morphosyntactic element guided infants to interpret the learned pseudo-word as describing an action and not an agent in an animated scene. English-learning infants were able to map trained pseudo-verbs (e.g., *larping*) to actions when pseudo-verbs contained a progressive morpheme *-ing* by two years of age (Waxman et al., 2009). Göksun, Küntay and Naigles (2008) found that a presence of an accusative morpheme reinforced an interpretation of Turkish verbs as causative in Turkish-speaking children starting at two years of age.

Gerken, Wilson and Lewis (2005) found that 16-month-old English-learning infants are sensitive to morphological markings when exposed to words from different syntactic categories of a natural language unfamiliar to them. In their study, infants were familiarized with a set of Russian words belonging to one of two gender categories. Each category occurred with two inflectional morphemes (e.g., xA1, xA2 for feminine gender, and yB1, yB2 for masculine gender, where x and y are word stems of feminine and masculine words, and A1, A2, B1, B2 are inflectional morphemes). After familiarization, infants were tested with Russian words which occurred only with one of the inflectional morphemes in the training (e.g., xA1). In grammatical test trials, the words appeared with the second morpheme of that gender category (e.g., xA2), whereas in ungrammatical test trials, they received a morpheme of the wrong gender category (e.g., xB1). Infants showed a looking preference for ungrammatical trials. These results suggest that inflectional morphemes help infants in learning of syntactic categories. In addition, the study raised a question of the role of double-markings. In some experimental conditions, a number of stimuli contained an additional marking of a suffix. Infants' preference for ungrammatical trials was stronger when both training and test stimuli contained some double-marked words. The effect was weaker when a majority of training stimuli were double-marked, while test stimuli were single-marked. When both training and test stimuli were unmarked, infants did not show any significant looking preference for grammatical or ungrammatical test trials.

Infants' knowledge of inflectional morphemes can be observed in the studies on non-adjacent dependencies which were reviewed in details in a previous section. Thus, Santelmann and Jusczyk (1998) showed that by 18 months of age, English-learning infants know that a progressive *-ing* morpheme requires a function word *is*. In this study, infants heard grammatical passages of sentences where the *is* + *ing* dependency was respected (e.g., *At the bakery, everybody* ***is baking*** *bread*). They also heard ungrammatical passages where the *-ing* morpheme incorrectly occurred after an auxiliary *can* (e.g., *At the bakery, everybody* ***can baking*** *bread*). Infants showed a looking preference for grammatical sentences. Höhle et al. (2006) found similar results with 19-month-old German-learning infants in present perfect constructions. Infants showed a looking preference for a grammatical dependency of a function word *haben (have)* with a past participle morpheme *-t*. The grammatical trials were contrasted with an ungrammatical dependency of the same morpheme with a function word *können* (*can*).

A study by Marquis and Shi (2012) directly tested the recognition of bound morphemes in French-learning eleven-month-olds. They trained infants with one of two pseudo-words mimicking a bare root of a verb (i.e., */trid/* for one group of infants and */glyt/* for another group). When the bare root was inflected with a frequent French */e/* morpheme in the test, infants showed a looking preference for the trained pseudo-word. When the bare root was inflected with a morpheme */u/* not existing in French, infants did not show any looking preference. These results suggest that by eleven months, French-learning infants recognize a frequent morpheme and can relate it to an uninflected form.

Marquis and Shi (2012) also inquired into the mechanism of infants' morphological learning. In a subsequent experiment, infants were trained with a set of pseudo-words, each containing the non-existent */u/* morpheme. After such training, they were able to link a novel pseudo-word containing the */u/* morpheme with an uninflected form of that word. As suggested by the authors, this learning occurred due to different frequencies of verb roots and the morpheme in the training. The verb roots were of highly variable types, with a low frequency of occurrence, whereas the artificial morpheme had only one type and a high frequency of occurrence. This study suggests that frequency effects play a role in morphological learning in preverbal infants.

Overall, the literature reviewed suggests that infants are sensitive to morphological markings. They rely on morphological markings when exposed to syntactic categories of an unknown language in the absence of semantic information (Gerken, Wilson and Lewis, 2005). Infants are sensitive to morphological markings prior to acquiring the semantics of their native language (Marquis & Shi, 2012). Morphological markings help infants in finding syntactic categories in their input (Gerken, Wilson and Lewis, 2005) and in relating inflected and non-inflected forms (Marquis & Shi, 2012).

1.6 Brief summary of the literature review

Generalization and the learning of exceptions are important features of language acquisition. The studies described in the literature review suggest that preverbal infants can generalize abstract rules to novel instances (Marcus et al., 1999; Gerken, 2006). The rules learned and generalized by infants in these studies were based on simple identity patterns (e.g., AAB). Our aim is to replicate and extend these findings to an artificial rule involving a movement of word order.

Previous research also showed that noise in the learning input can impede infants' generalization (Gómez and LaKusta, 2004). Low levels of noise can still be tolerated, but when it increases, infants' generalization fails. It is not clear, however, whether this effect of noise is due to overall noise frequency or to noise type frequency. No study has directly tested how the relative frequency of noise and rule-based exemplars in the input affect infants' generalization.

Sometimes input can allow more than one generalization. Gerken (2006) examined whether infants make a larger or more conservative generalization when they encounter such input. It was found that infants make a more conservative generalization that closely follows the properties of the input. In that study, infants were learning a rule from perfectly consistent input where all exemplars conformed to a rule. What kind of a generalization infants make when the input contains noise has not previously examined.

Noise can be examplars to which the rule does not apply. An important question about learnability is how children know whether such non-application cases are true exceptions, or just regular cases that they had never previously heard with the rule. Production studies show that children sometimes make over-generalization errors by applying a rule to true exceptions (see a review in Bowerman, 1988). Frequency seems to play a role in children's learning of exceptions, as suggested by the notion of Entrenchment (Braine & Brooks, 1995), according to which, when a word occurs frequently in a particular construction, it is less likely to be over-generalized to a wrong construction. The role of frequency in over-generalization and the learning of exceptions was supported by empirical studies. Brooks et al. (1999) observed this effect with children who made fewer over-generalization errors with high-frequency English verbs. In other studies, children and adults tolerated over-generalization errors more for low frequency verbs (Theakston, 2004; Ambridge et al., 2008). Another study with adults suggested that token frequency plays an important role in whether adults learn the trained verbs as exceptions entrenched to a minority structure (Wonnacott, Newport and Tanenhaus, 2008). How token frequency can affect over-generalization and the learning of exceptions in infants has not been previously tested.

1.7 Research questions and hypotheses

In this work we address the questions of abstract rule learning and generalization in infants. Do type and token frequency affect infants' generalization of a learned rule to novel instances? Under what conditions do infants make over-generalizations of instances that do not conform to the rule? How do infants resist over-generalization and learn exceptions?

We tested this by creating two artificial rules of word order movement, using a natural language (Russian) unfamiliar to our infant participants. In one rule, sentences with an ABC order moved immediately into the BAC order (e.g., *Chistim tufli vaksoj – Tufli chistim vaksoj*). The word order pairs were always immediately adjacent. In another rule sentences with an ABC order moved immediately into the ACB order (e.g., *Chistim tufli vaksoj – Chistim vaksoj tufli*). Infants were trained with sentences conforming to one of these rules. In some experiments, the training input also contained additional exemplars with an ABC structure that served as noise (e.g., *Otzvuk smekha sladok*). The noise instances did not go through either movement rule. The noise sentence could occur between two rule pairs, between a rule pair and another noise sentence or between two noise sentences. The abstraction and generalization of rules were tested with novel sentences going through the trained and untrained rules (in Experiments 1 – 12). Over-generalization and the learning of exceptions were tested with noise instances from the training that followed the trained and untrained rules (in Experiments 13 – 15).

Based on the previous work of Marcus et al. (1999), we expect that infants can learn an abstract rule from 100% consistent input and generalize it to novel instances (in Experiments 1 and 2). We also predict that infants will tolerate a small amount of noise in their learning input (in Experiments 3 – 10), as in the study by Gómez and LaKusta (2004). We further predict that in order to allow abstraction and generalization of rules the amount of noise needs to be low in type frequency.

Concerning over-generalization and the learning of exceptions, we expect that they are affected by the token frequency of noise exemplars (in Experiments 13 – 15).

We predict that a low token frequency of noise instances leads to over-generalization whereas high token frequency of noise instances leads to the learning of exceptions.

We also examine the nature of the rule learned by infants when two generalizations are possible (in Experiments 11 and 12). We expect that when larger and more conservative generalizations are possible, infants make the more conservative generalization, the one closer to the properties of the input. This prediction is based on findings from a study by Gerken (2006).

CHAPTER II

EXPERIMENTS

2.1 Rule learning and generalization from 100% consistent input

2.1.1 Experiment 1: Training – 100% consistent; test – novel instances; no morphological markings; age – 11-month-olds

The purpose of this series of experiments was to analyze the role of rule and noise distributions in infants' learning and generalization. The first experiment tested the age at which infants can learn and generalize to novel instances an abstract movement rule in an unfamiliar natural language when the input consistently supports the rule. Although there already exists some research on infants' learning and generalization in conditions of perfectly consistent input, the results cannot be directly applied to our experimental conditions because of the unequal complexity of the stimuli. The stimuli of Marcus et al. (1999) and Gerken (2006) were simple, monosyllabic CV words combined into three-word strings by rules, and each string included a reduplication of one of two elements (e.g., AAB). Infants as young as seven months of age learned the rules in Marcus et al. (1999), although Gerken (2006) did not fully replicate those findings with seven-month-olds. Gerken obtained robust results only with nine-month-old infants. Since our interest was in the learning of a natural language, we used a language unfamiliar to infants (Russian). Experiment 1 used highly variable multisyllabic Russian words, which were stimuli of greater complexity than in the studies by Marcus et al. (1999) and Gerken (2006). The rules we used were abstract word order movement rules (i.e., ABC –> BAC and ABC –> ACB). The word order movement rules were more complex than the identity-based rules used by Marcus et al. (1999) and Gerken (2006), as infants were required to register each sentence in memory and to track its moved version. Given that our task was more demanding overall than that of Marcus et al. (1999) and Gerken (2006), we chose an older age group: 11-month-old infants.

2.1.1.1 Participants and Materials

Sixteen infants aged 11 months from various linguistic backgrounds completed the experiment. The age of the 9 boys and 7 girls ranged from 11 months 3 days to 11 months 27 days (M = 11 months 10 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Four other infants were tested but their data were not included in the analysis for various reasons such as parental interference (1), experimenter error (2) and ceiling effect (i.e., looking for the maximal trial length during all test trials) (1). Two other infants did not complete the experiment.

Materials were 12 Russian sentences (see pp. 148-149, Appendices A and B) recorded by a female native Russian speaker in the child-directed speech style. The speaker clearly separated the words when producing each sentence. Eight of the sentences were used as training stimuli, and four as novel instances in the test phase. All original ABC sentences had a Subject-Verb-Subordinate structure (i.e., S-V-Sbd). The subordinate part of sentence was variable. It could be an adverbial, a direct object or an indirect object. This design served to diversify the morphological features. All the words were highly variable in phonotactic and morphological properties, and in the number of syllables.

Training input consisted of eight sentences, each occurring four times. Training sentences presented a word order movement rule that was applied to each sentence. For example, a sentence with ABC word order was followed immediately by the same sentence transformed into BAC for one training condition (Rule 1), or into ACB for the other condition (Rule 2). In order to ensure variability, various exemplars from the recording were used for the eight training sentences and their inversions: two exemplars for four sentences and their inversions, three exemplars for three sentences and their inversions, and four exemplars for one sentence and its inversions. Thus, the exemplar distribution was equal for both rules. Four strings were composed for each rule, each containing a random sequence of eight original sentences and their moved versions. The moved version followed immediately after each original sentence.

Test materials consisted of four novel sentences not included in the training sets. One recording was used for each test sentence. Two of the four test sentences underwent the "ABC to BAC" movement (Rule 1); the other two novel sentences underwent "ABC to ACB" (Rule 2).

Sentences within each sentence pair (i.e., the original and its moved version) were separated by approximately 700 ms, in both training and test. Sentence pairs were separated from other sentence pairs by approximately 1200 ms. In the training, average sentence duration was 3.86 s ($SD = 0.47$) for Rule 1 and 3.73 s ($SD = 0.48$) for Rule 2. In the test, average sentence duration was 4.24 s ($SD = 0.19$) for Rule 1 and 4.19 s ($SD = 0.28$) for Rule 2.

The visual stimulus for all trials was an animation with many multi-colored circles of varying size. A sinusoidal sound was used for contingency training and for the post-experiment trial. An animation of blue bubbles accompanied by a cricket sound served as the attention-getter.

2.1.1.2 Design and Procedure

The experiment consisted of the following steps:

1. Training: passive listening phase, in which each infant was exposed to either the Rule 1 or the Rule 2 training set (i.e., "ABC to BAC" or "ABC to ACB" sentences). The total duration of the training phase was 314 s for Rule 1 and 305 s for Rule 2.

2. Pre-test: each infant heard one trial with two novel test sentences undergoing the ABC to BAC (i.e., S-V-Sbd to V-S-Sbd) transformation, and another trial with two other novel sentences undergoing ABC to ACB (i.e., S-V-Sbd to S-Sbd-V). The trials were identical to the test trials in Step 4 (see below). Each trial was initiated when the infant looked at the center screen. This pre-test phase allowed infants to hear one full version of each test stimulus regardless of whether they continued to look at the screen. This served as a basis for the potential recognition of particular sentences associated with one of the two rules after infants began hearing the early part of the stimuli in the test phase (Step 4).

The order of the two sentences within a trial was fixed. The order of the two types of test trials (Rule 1 vs. Rule 2) was counterbalanced across infants: some infants heard test sentences conforming to Rule 1 first, others heard Rule 2 first. Each pre-test trial had a fixed length of 21 sec.

3. Contingency training: two contingency training trials were designed to teach the infants that they could fully control the duration of trials. Auditory stimuli were sinewave sounds. A trial was initiated when the infant looked at the screen, and it terminated if the infant looked away from the screen. Minimum look-away for terminating a trial was 2 s. The maximum duration of each trial was 9 sec if looking

lasted till the end of a trial. Trials starting from this step were all fully infant-controlled.

4. Test phase: 10 test trials. This phase presented exactly what infants had heard during the pre-test trials (Step 2), except that the trials were fully infant-controlled. Maximum trial duration was 21 s if the infant looked till the end of a trial. The counterbalancing of the order of the two trials was in line with that of Step 2. For example, if Rule 1 was presented first in Step 2, it was also presented first in Step 4.

5. Post-experimental phase: one trial identical to the contingency training trials, except that the maximum trial length was 21 s. This trial enabled us to determine whether the infants were on task throughout the experiment. If so, the looking time should increase during this post-trial relative to the last test trial, as the auditory examples were distinct from those of the 10 test trials.

Each infant was tested individually. For the passive listening phase of training (Step 1), the infant and the parent were invited to a sound chamber. There was a TV screen and a sofa in the room, and speakers for auditory presentation were next to the left and right sides of the TV. The infant was given toys, a way to forestall boredom. Parents were instructed to keep silent. They could play silently with the child. The child could move around freely in the room. During the presentation of the sentences, the infants saw an animation on the screen, with bright multi-colored circles slowly changing sizes.

After the passive listening phase, the parent and infant left the toys behind and moved to another acoustic chamber for Step 2-5 of the experiment, which were executed by an experimental program (Cohen, Atkinson, & Chaput, 2000). The infant sat on the parent's lap facing the TV monitor. The parent wore headphones to hear masking music. She or he was asked not to interfere with the infant's reactions. The experimenter, who was blind to the audio-visual stimuli, observed the infant's eye movements from a closed-circuit TV in an adjacent room, and pressed down a computer key whenever the infant looked at the screen. The experimental software presented the stimuli and automatically recorded all looking times. Each trial in Steps 2-5 was initiated by the infant's looking toward the screen.

Half of the infants were trained with Rule 1, the other half with Rule 2 (Step 1). In the test phase (Step 4), all infants were presented with two Rule 1 and two Rule 2 novel sentences, the same stimuli as in Step 2, in separate trials. There were two test trial types. In one type the two sentences conformed to the trained rule, whereas in another type the other two sentences conformed to the non-trained rule. These trial types were presented in alternation for 10 test trials in all (5 for each type). As part of the counter-balancing, half the infants were presented with novel test sentences that conformed to the trained rule as the first test trial, and the other half heard sentences that conformed to the non-trained rule as the first test trial.

For all trials, the auditory stimuli were presented simultaneously with visual stimuli consisting of circles growing and reducing in size on the screen. Between trials, the attention-getter (blue bubbles accompanied by a cricket sound) was presented.

2.1.1.3 Results

Each infant's looking times during the two trial types in the test phase (Step 4) were calculated, i.e., the trials presenting sentences conforming to the trained rule and those presenting sentences conforming to the other rule that infants did not hear during training. A Paired Samples t-test failed to show any significant discrimination of these two trial types, $t(15) = -0.84$, $p = 0.412$, *two-tailed, partial eta squared* = 0.045. Cumulative looking time across test trials was on average 38.56 s ($SE = 4.42$)

for the trained movement rule and 42.05 s (SE = 4.41) for the non-trained movement rule. Average looking time per trial was 7.71 s (SE = 0.88) for the trained and 8.41 s (SE = 0.88) for the non-trained rule. (In this and further experiments of this thesis a significance level was set at an alpha probability of 0.05.)

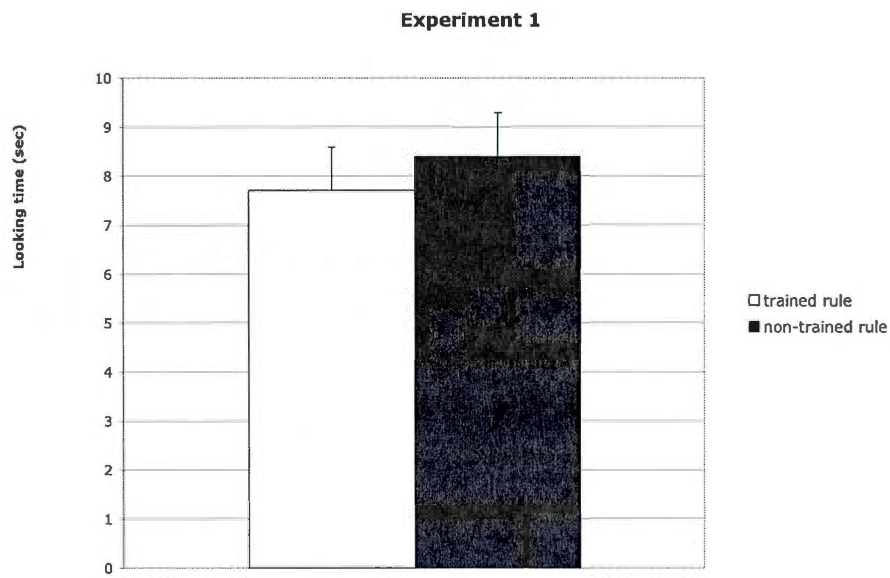

Figure 2.1 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 1: Training – 100% consistent; test – novel instances; no morphological markings; age – 11-month-olds. Infants' looking times for the two types of test trials were not significantly different.

There was hence no evidence that 11-month-olds had the capacity to generalize abstract movement rules after a brief exposure to an unfamiliar language. In Experiment 2 we examined whether older infants (14-month-olds) could learn abstract movement rules from natural language input.

2.1.2 Experiment 2: Training – 100% consistent; test – novel instances; no morphological markings; age – 14-month-olds

2.1.2.1. Participants and Materials

Sixteen infants (10 boys and 6 girls) aged 14 months from various linguistic backgrounds completed the experiment. The age ranged from 13 months 30 days to 14 months 26 days (*M* = 14 months 15 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had had any prior exposure to Russian. Six other infants were tested, but their data were not included in the analysis for various reasons such as moving out of camera range during test trials (2), parental interference (1), experimenter error (1), singing during test trials (1) and looking at the screen for only 2 seconds or less on 6 or more test trials (1). One other infant did not complete the experiment.

Materials were the same as in Experiment 1.

2.1.2.2 Design and Procedure

Design and procedure were identical to those of Experiment 1.

2.1.2.3 Results

Each infant's looking times during the two trial types in the test phase (Step 4) were calculated, i.e., the trials presenting sentences conforming to the trained rule and those presenting sentences conforming to the other rule that infants did not hear during training. A Paired Samples *t*-test revealed that infants showed a significant discrimination of these two rules, $t(15) = -2.44$, $p = 0.027$, *two-tailed, partial eta* *squared* = 0.284. Cumulative looking time across test trials was on average 26.75 s (*SE* = 3.56) for the trained movement rule and 36.12 s (*SE* = 3.47) for the non-trained movement rule. Average looking time per trial was 5.35 s (*SE* = 0.71) for the trained movement rule and 7.22 s (*SE* = 0.69) for the non-trained movement rule.

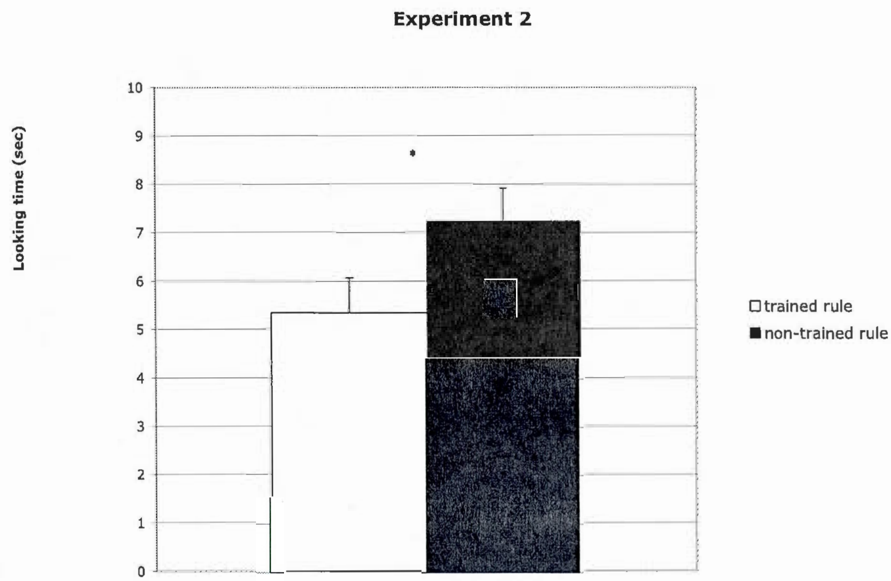

Figure 2.2 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 2: Training – 100% consistent; test – novel instances; no morphological markings; age – 14-month-olds. Infants' looking times for the two types of test trials were significantly different.

The results suggest that after a brief exposure to an unfamiliar natural language, 14–month-olds can learn movement rules and generalize them to novel instances, in the absence of any phonological, morphological and semantic cues. In all subsequent experiments of this thesis, only 14-month-old infants were tested.

2.2 The role of type frequencies in infants' learning and generalization of movement rules

2.2.1 Experiment 3: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; morphologically marked

In Experiments 1 and 2 we determined that 14-month-old infants were able to learn movement rules and generalize them to novel instances after a brief exposure to an unfamiliar natural language. The training input was 100% consistent with a rule in those experiments. In Experiment 3 and all subsequent experiments, we further examined how infants generalize abstract rules when exposed to less consistent input containing noise sentences, i.e., sentences not following the rule.

One factor to examine was the type variability of rule versus noise instances in the input. Noise instances that do not conform to a rule can be many things. They can either violate a rule explicitly (e.g., sentences directly showing a different pattern), or they can do so implicitly, by not applying a rule (which is not an overt violation of the rule). For example, a sentence with an ABC structure does not overtly violate the ABC -> BAC rule. Cases of non-application can potentially be perceived by the learner as being possible, would-be rule cases. In previous research on infants' learning and generalization from inconsistent input, the noise instances were cases of overt violation of the rule, as in the study by Gómez and LaKusta (2004): two word categories were combined inappropriately. They found that infants failed to learn the rule from the input in which the frequency of noise instances reached 33% relative to the frequency of the ruleful examples. Our interest, however, was in cases of non-application. Given the abstract ABC to BAC movement rule, non-application would be simple ABC sentences not going through any movement. Such noise cases can be interpreted as 'true' noise, i.e. cases which should never go through movement. They can also be interpreted as cases which did not yet have the chance to go through the movement, but the application of the rule is possible and grammatical. These two possible interpretations can be compared to children's learning of the irregular verb '*hit*' in English. Infants never hear '*hitted*'. That is, an overt rule violation for this word never occurs in the input. Likewise, in our experiments the noise instances did not overtly violate the rule. The question was how infants interpret non-applications at the early stages of language acquisition when semantic information is not yet accessible. In our design, rule-conforming sentences were the ones where ABC sentences went immediately through either BAC, or ACB movement. For example, *Vika darit murku* would immediately change into *Darit Vika murku.* This is an example of rule application. Noise sentences in our design were non-application cases. For example, a sentence with an ABC structure did not change into either BAC or ACB sentences and stayed unchanged: e.g., *Gena vidit lavku.*

Experiment 3 tested how type frequency of rule and noise cases in the input can affect infants' capacity to abstract rules and apply them to novel instances. To do so, we dissociated type and token frequencies of rule and noise instances in the training sample. The training consisted of eight rule-based sentences. Each of them went through the movement transformation. In addition, there were two noise sentences. Neither of them moved. Therefore, the type frequency was 80% for ruleful examples and 20% for noise. Rule instances each occurred four times. That is, the token frequency per type was four, i.e., the type-token ratio was 1:4. The noise sentences each occurred 16 times, i.e., the type-token ratio was 1:16. Hence the overall token frequency was equal for rule examples (8x4=32) and noise examples (2x16=32).

If the relative type frequency of rule instances is the determining factor for rule learning, infants should succeed in generalizing the trained rule to novel instances. But if the token frequency per type is the determining factor, infants should fail to learn the rule. If the overall frequency of rule instances must be higher than that of noise, infants should also fail to learn the rule.

2.2.1.1 Participants and Materials

Sixteen infants (8 boys and 8 girls) aged 14 months from various linguistic backgrounds completed the experiment. Their ages ranged from 14 months 15 days to 15 months 05 days (*M* = 14 months 26 days). Parents were asked about their children's language background (see p. 161, Appendix N). None had had any prior exposure to Russian. Eight other infants were tested but their data were not included in the analysis for various reasons such as fussiness (1), crying (2), parental interference (2), experimenter error (2), and looking toward the screen for 2 seconds or less on 6 or more test trials (1).

Materials were 12 new Russian sentences (see pp. 150-152, Appendices C, D and E) recorded by the same Russian native speaker in the same way as in Experiments 1 and 2. Ten of these sentences were used as training stimuli (see pp. 151-152, Appendices D and E), and two as novel instances in the test phase (see p. 150, Appendix C). Sentences had a Subject-Verb-Object structure. All words contained two syllables. In this experiment we selected Russian words which had consistent morphological markings for each position within the ABC sentences. All words in the A position ended with *–a*, those in the B position ended with *–it*, and those in the C position ended with *–ku*. The words in the moved sentences kept their original morphological markings.

Out of the ten training instances, eight were used as rule instances and two were noise instances. For Rule 1, each ABC sentence was immediately followed by its BAC version, whereas for Rule 2, each ABC was immediately followed by ACB. The two noise sentences went through no movement.

One recorded examplar was used for each original rule sentence (ABC) and each of its inverted versions (BAC, ACB). For each of the two noise sentences (ABC only), eight recorded examples were used.

There were two training sets, one containing the noise and Rule 1 cases, the other the same noise and Rule 2 cases. For each set, the eight rule sentence pairs (one recording exemplar each, i.e., 8x1) and the two noise sentences (four recording exemplars each, i.e., 2x4) were randomly arranged to form the first string. The second string was made by randomizing the same exemplars of rule pairs and replacing the noise exemplars of the first string with new recorded examplars (four for each noise sentence, i.e., 2x4). The third string was formed by reversing the first and second halves of the stimuli of the first string. Likewise, the fourth string was formed by reversing the two halves of the stimuli in the second string. These manipulations were done to create sufficient order variability during training. Within a sentence pair, the original and the moved version were separated by approximately 700 ms. The pause between rule and noise types, between any pairs, and between any two noise sentences was approximately 1200 ms. Therefore, four strings were prepared for each training input set.

Across the four strings of a training set, the eight rule pairs occurred four times each, and the two noise sentences occurred 16 times each. That is, the total number of occurrences of rule pairs and noise sentences, i.e., overall token frequency, was kept equal (8x4=32 rule instances, and 2x16= 32 noise instances).

The test stimuli were two novel ABC sentences and their moved versions (ABC-BAC for Rule 1 and ABC-ACB for Rule 2). Unlike in Experiments 1 and 2, where a test sentence moved only according to one of two rules, here both movements were created for each sentence. One recorded examplar was used for each original ABC sentence and for its inverted BAC and ACB versions. However, not all of them were presented to the same infant, as described in the next section. Sentences in the test phase were separated by the same inter-stimulus intervals (ISIs) as in the training phase.

In the training, average sentence duration was 2.49 s (*SD* = 0.2) for Rule 1, 2.51 s (*SD* = 0.21) for Rule 2, and 2.32 s (*SD* = 0.11) for noise sentences. In the test, average sentence duration was 2.53 s (*SD* = 0.096) for Rule 1 and 2.54 s (*SD* = 0.05) for Rule 2.

2.2.1.2 Design and Procedure

Design and procedure were nearly identical to Experiments 1 and 2. Here, total duration of the training strings was 341 s for the 'Rule 1 + Noise' training condition and 340 s for the 'Rule 2 + Noise' training condition. Another important difference from Experiments 1 & 2 was that Step 2 and Step 4 differed by the number of sentences in the trial and by the maximum trial length. Unlike in Experiments 1 and 2, where each pre-test and test trial contained two different sentences and their moved versions, in Experiment 3 each pre-test and test trial contained only one sentence and its moved version. Here, in the pre-test trials (Step 2), the sentence pair occurred only once, whereas in test trials (Step 4), the same sentence pair was presented up to three times if the infant looked till the end of the trial. The duration of each pre-test trial was of fixed length 6 s (Step 2), and the maximum trial length of each test trial was 20 s (Step 4).

As in Experiments 1 and 2, the test phase was characterized by two types of trials, one type for the trained rule, and the other type for the non-trained rule. In one test trial, one of the two novel test sentences went through the ABC-BAC rule. In the other test trial, the second sentence went through the ABC-ACB rule. The sentence and rule application were counter-balanced across infants, e.g., one group of infants heard the first sentence as ABC-BAC and the second sentence as ABC-ACB, while another group heard the first sentence as ABC-ACB and the second as ABC-ACB. The order of test trials was also counter-balanced, so half the infants heard the ABC-BAC rule as the first test trial whereas the other half heard the ABC-ACB rule as the first test trial. The two types of trials alternated, yielding a total of 10 test trials.

As in Experiments 1 and 2, the dependent measure was the looking time of infants towards the screen while listening to test trials conforming to the trained rule, and test trials following the untrained rule. If infants can learn and generalize the movement rule from the noisy input, then we should obtain significant looking time differences for the two trial types in the test phase.

2.2.1.3 Results

As in Experiments 1 and 2, each infant's looking times during the two test trial types were calculated. A Paired Samples t-test revealed that looking times for the two trial types were significantly different, $t(15) = -2.65$, $p = 0.018$, *two-tailed, partial eta squared* = 0.318. Average cumulative looking time across test trials was 29.46 s (SE = 3.63) for the trained rule and 42.34 s (SE = 5.37) for the non-trained one. Average looking time per trial was 5.89 s (SE = 0.73) for the trained rule, 8.47 s (SE = 1.07) for non-trained.

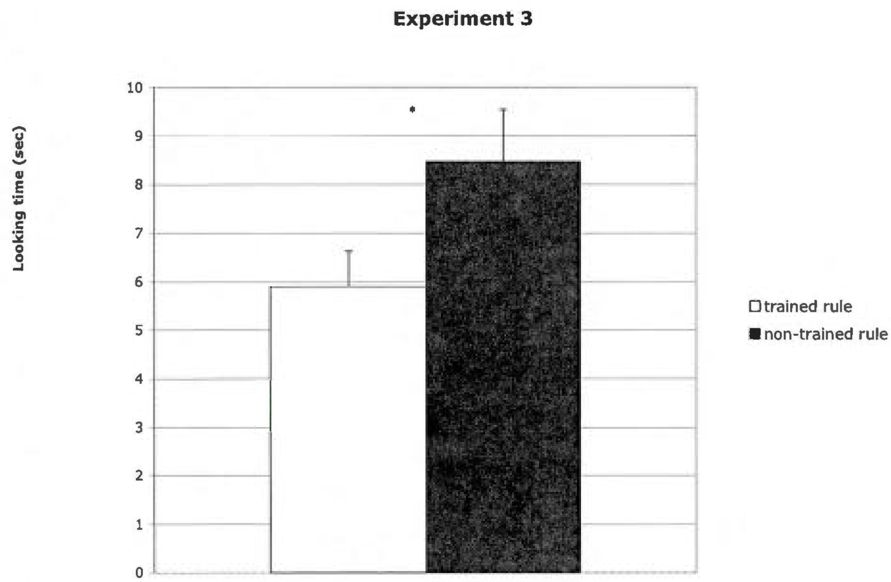

Figure 2.3 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 3: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; morphologically marked. Infants' looking times for the two types of test trials were significantly different.

The results suggest that after a brief exposure to an unfamiliar natural language, 14–month-olds can learn movement rules and generalize the rules to novel instances. Crucially, our results showed that dominant type frequency of rule instances allowed infants to learn and generalize the rule in the conditions of the inconsistent learning input. However, is it possible that infants simply did not treat the non-application instances as noise? If so, they should succeed in the learning regardless of the proportion of noise sentences. Experiment 4 addressed this question.

2.2.2 Experiment 4: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; morphologically marked

The purpose of this experiment was to control for the results of Experiment 3 and examine whether learning could take place when the type frequency of rule instances in the training was not dominant. In order to do so, we increased the type frequency of noise sentences from two to eight while presenting exactly the same number of rule instances and their occurrences (8x4) as in Experiment 3. Each rule sentence went through movement. No noise sentences moved. Type frequency was 50% for rule and 50% for noise. Each rule type occurred four times, i.e. the type-token ratio was 1:4. Each noise sentence also occurred four times, i.e., the type-token ratio was 1:4. In other words, the token frequency per type was four. The overall frequency was the same for rule and noise (8x4=32). Hence, rule and noise sentences were equal by type, type-token ratio and overall frequency.

In all other respects, the design was similar to Experiment 3. Again, the measure was infants' looking time toward the screen while listening to test trials with the trained rule and test trials with the untrained rule. If type frequency does play a crucial role in the learning of abstract rujles, as suggested by Experiment 3, then infants should not show any learning in Experiment 4.

2.2.2.1. Participants and Materials

Sixteen infants aged 14 months from various linguistic backgrounds completed the experiment. The age of the 9 boys and 7 girls ranged from 14 months 3 days to 14 months 24 days (M = 14 months 15 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Fourteen other infants were tested but their data were not analysed for various reasons such as fussiness (5), getting out of camera range during test trials (1), crying (1), lack of interest (1), parental interference (3), and looking at the screen for 2 seconds or less in 6 or more test trials (3). One other infant did not complete the experiment.

The sentences from Experiment 3 were used for Experiment 4, and six new noise sentences were added, yielding eight rule instances (see p. 151, Appendix D) and eight noise instances (see p. 153, Appendix F) for each training set. The new noise sentences were recorded by the same speaker who produced the stimuli of Experiments 1-3. Four training strings were constructed for each of the two training sets. In the first string, the eight rule pairs and eight noise sentences occurred once each in random order. The second string contained the same eight rule pairs and eight noise instances, in a differently randomized order. A different recording exemplar was used for each of the two old noise sentences from Experiment 3, but all remaining rule and noise exemplars for the second string were those from the first string. The third and fourth strings were made by reversing the early and late halves of the first and second strings, respectively. ISI was the same as in Experiment 3. Average sentence duration for all noise examples was 2.39 s (*SD* = 0.15). Test examples for Experiment 3 were the same as in Experiment 4.

2.2.2.2 Design and Procedure

Design and procedure were nearly identical to Experiment 3, except that eight noise sentences were presented four times each (mixed with eight rule pairs presented four times each). Total duration of the training phase was 341 s for the 'Rule 1 + Noise' training condition and 340 s for the 'Rule 2 + Noise' training condition.

2.2.2.3 Results

Each infant's looking times during the two trial types were calculated, i.e., the looking times during the sentence conforming to the trained rule versus that conforming to the other rule which had not been present in the training. A Paired Samples *t*-test revealed that infants showed no significant discrimination between these two rules, $t(15) = 0.18$, $p = 0.858$, *two-tailed, partial eta squared* = 0.002. Cumulative looking time across test trials was on average 36.99 s (*SE* = 5.29) for the trained movement rule and 35.96 s (*SE* = 4.52) for the non-trained movement rule. Average looking time per trial was 7.4 s (*SE* = 1.06) for the trained rule and 7.19 s (*SE* = 0.91) for the non-trained.

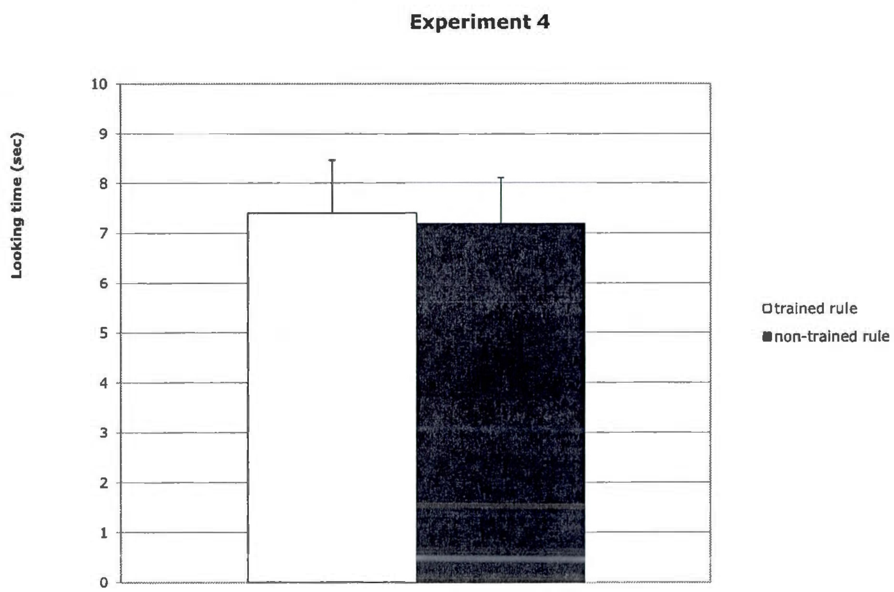

Figure 2.4 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 4: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; morphologically marked. Infants' looking times for the two types of test trials were not significantly different.

These results suggest that when rule instances were no longer dominant by type frequency, infants could no longer learn and generalize the abstract movement rule. We note that rule sentences in Experiment 4 were presented with exactly the same number of types and tokens as in Experiment 3. The crucial difference was the proportion of types for rule instances relative to noise: in Experiment 3, rule instances were dominant by type frequency, and infants showed learning, whereas in Experiment 4, rule instances were no longer dominant by type frequency due to the increase of type frequency of noise, and learning was impeded. The combined results of Experiment 3 and 4 suggest that rule type frequency is important for rule learning and generalization. In Experiments 5 and 6, we tested whether morphological markings were required for this learning.

2.2.3 Experiment 5: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; no morphological markings

In Experiment 3, where infants showed learning, training and test stimuli had consistent morphological markings for each word in the A, B and C position. Morphological markings may have assisted rule learning in that experiment. To examine this interpretation, we designed Experiment 5, which was identical to Experiment 3 in all respects except that the morphological markings in the familiarization and test stimuli were inconsistent. For infants unfamiliar with Russian, such stimuli did not indicate any morphological cue. Therefore, in all the following description, the absence of consistent morphological markings will be identified as 'no markings'.

As in Experiment 3, the training consisted of eight rule instances. Each immediately went through the movement transformation. In addition, there were two noise sentences. Neither of them moved. Hence the type frequency was 80% for rule and 20% for noise. Each rule type occurred four times. That is, the token frequency per type was four, i.e. the type-token ratio was 1:4. Each noise sentence occurred four times more, i.e., the type-token ratio was 1:16.

If morphological markings are not required for rule learning from input with dominant type frequency of rule instances then infants should discriminate between the trained and the non-trained rules being applied to novel morphologically unmarked instances. If, on the contrary, morphological markings are required, then infants should show no discrimination in this experiment.

2.2.3.1 Participants and Materials

Sixteen infants (10 boys and 6 girls) aged 14 months from various linguistic backgrounds completed the experiment. The age ranged from 14 months 9 days to 14 months 29 days (*M* = 14 months 18 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Eight other infants were tested but their data were not included in the analysis for various reasons such as fussiness (1), crying (2), parental interference (3), experimenter's error (1), and looking to the screen for 2 seconds and less during 6 or more test trials (1). Four other infants did not complete the experiment.

The same Russian native speaker recorded 12 new sentences (see pp. 147-149, Appendices G, H and I). Ten of these sentences were used as training stimuli (see pp. 154-156, Appendices H and I), and two as novel instances in the test phase (see p. 154, Appendix G). All words contained two syllables. To obtain the unmarked learning input, the original sentences (ABC) had variable parts of speech for words in A, B and C positions. For example, words in the A position could be a noun, a verb or an adverb.

As in Experiment 3, eight sentences were used as rule instances (see p. 155, Appendix H) in the training. Two other sentences were used as noise (see p. 156, Appendix I). Each of the rule instances went immediately through the movement transformations (Rule 1: ABC -> BAC; Rule 2: ABC -> ACB). The two noise sentences went through no movement.

Similarly to Experiment 3, we used one recording for each original rule sentence (ABC) and for each of its inverted versions (BAC, ACB). Unlike in Experiment 3, we used four (and not eight) recordings for each of the two noise sentences.

As in Experiment 3, one training set contained cases of Rule 1 and noise, and the other Rule 2 plus the same noise examples. For each set, the eight rule sentence pairs (one recording exemplar each, i.e., 8x1) and the two noise sentences (four recording exemplars each, i.e., 2x4) were randomly arranged to form the first string. The second string was made by randomizing the same exemplars of rule pairs and noise sentences, except that one recording exemplar of the noise sentence was used twice in the second string. The second occurrence of that recording exemplar within the second string replaced a recording exemplar from the first string. The third string was formed by reversing the first and the second halves of the first string. Likewise, the fourth string was formed by reversing the two halves of the second string. Within a sentence pair, the original and the moved version were separated by approximately 700 ms. The pause between rule and noise types, between any pairs, and between any two noise sentences was approximately 1200 ms. Hence four strings were prepared for each training input set.

In the training, average sentence duration was 2.63 s (SD = 0.19) for Rule 1, 2.65 s (SD = 0.18) for Rule 2, and 2.59 s (SD = 0.16) for noise sentences. In the test, average sentences duration was 2.55 s (SD = 0.09) for Rule 1 and 2.55 s (SD = 0.09) for Rule 2.

2.2.3.2 Design and Procedure

Design and procedure were identical to Experiment 3. Total duration of the training phase was 349.7 s for the 'Rule 1 + Noise' training condition and 350.3 s for 'Rule 2 + Noise'.

2.2.3.3 Results

Each infant's looking times during the two test trial types (trained rule versus non-trained rule) were calculated. A Paired Samples t-test revealed that infants showed a significant discrimination between these two rules, $t(15) = -2.82$, p = 0.013, *two-tailed, partial eta squared* = 0.347. Cumulative looking time across test trials was on average 41.19 s (SE = 6.01) for the trained movement rule and 49.4 s (SE = 6.32) for the non-trained movement rule. Average looking time per trial was 8.24 s (SE = 1.2) for the trained movement rule and 9.88 s (SE = 1.27) for the non-trained movement rule.

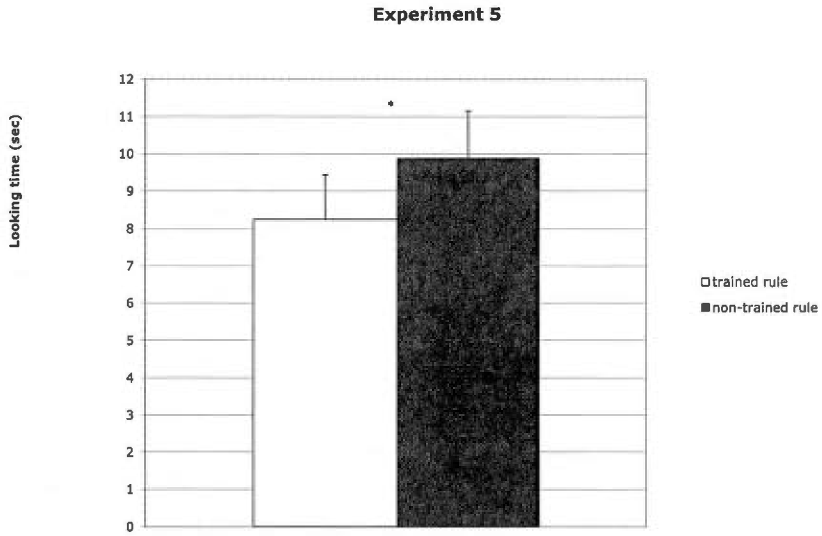

Figure 2.5 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 5: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; no morphological markings. Infants' looking times for the two types of test trials were significantly different.

These results suggest that infants successfully learned the trained movement rule and generalized it to novel instances even in the absence of morphological markings. In this experiment, their learning was based solely on the dominant type frequency of rule-conforming instances. The results show that morphological markings are not required for infants to learn a movement rule on the basis of type frequency. To further confirm this interpretation, we conducted Experiment 6, in which rule instances in the training input were no longer dominant by type frequency. In that sense, this experiment had the identical rule and noise distribution as Experiment 4. Since the new experiment served to control for the results obtained in Experiment 5, we again used stimuli without morphological markings.

2.2.4 Experiment 6: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; no morphological markings

As in Experiment 4 the training consisted of eight rule instances and eight noise instances. Each rule and noise instance occurred four times, i.e. the type-token ratio was 1:4 for both. Hence, rule and noise instances were equal by type, type-token ratio and overall frequency, Unlike in Experiment 4, the stimuli contained no morphological markings. If the successful learning in Experiment 5 was due to the dominant type frequency of rule instances, then infants should fail to learn in Experiment 6.

2.2.4.1 Participants and Materials

Sixteen infants from various linguistic backgrounds completed the experiment. The parental report of children's language backgrounds is given in the Appendix N (p. 161). None of the infants had any prior exposure to Russian. The age of the 9 boys and 7 girls ranged from 14 months 13 days to 15 months 14 days (M = 15 months 00 days). Seven other infants were tested but their data not included in the analysis for various reasons such as fussiness (1), crying (1), experimenter's error (1), and looking to the screen for 2 seconds and less during 6 or more test trials (4). Five other infants did not complete the experiment.

The sentences in Experiment 5 were used for Experiment 6, and six new noise sentences were added, yielding eight rule instances (see p. 155, Appendix H) and eight noise instances for the training (see p. 157, Appendix J). The new noise sentences were recorded by the same speaker who produced the stimuli for the previous experiments. Four training strings were constructed for each of the two training conditions. Within the first string, the eight rule pairs and eight noise instances occurred once each in a random order. The same recording exemplars of rule and noise instances in the first string were randomly rearranged to form the second string. The third and fourth strings were made by reversing the two halves of the stimuli within the first and second strings, as in the previous experiments. ISIs were the same as in Experiment 5. Average sentence duration of all noise sentences was 2.46 s ($SD = 0.18$). Other sentences were identical to those in Experiment 5.

Test stimuli were the same as in Experiment 5.

2.2.4.2 Design and Procedure

Design and procedure were nearly identical to those in Experiment 5, except that here, the distribution of rule and noise in the training was different. In this respect, the training was identical to Experiment 4 (both type, token per type, and overall frequencies were at 50%). The total duration of the training phase was 343.7 s for the 'Rule 1 + Noise' training condition and 344.4 s for 'Rule 2 + Noise'.

2.2.4.3 Results

Each infant's looking times during the two test trial types (trained rule versus non-trained rule) were calculated. A Paired Samples *t*-test revealed no discrimination between these two rules, $t(15) = 0.56$, $p = 0.585$, *two-tailed, partial eta squared* = 0.02. Cumulative looking time across test trials was on average 39.89 s ($SE = 6.63$) for the trained movement rule and 37.56 s ($SE = 5.18$) for the non-trained movement rule. Average looking time per trial was 7.98 s ($SE = 1.33$) for the trained movement rule and 7.51 s ($SE = 1.04$) for the non-trained movement rule.

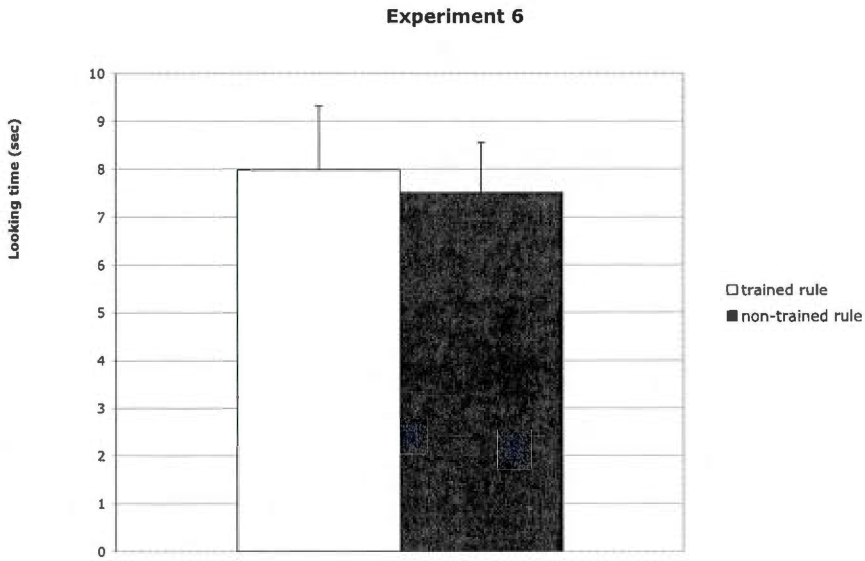

Figure 2.6 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 6: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; no morphological markings. Infants' looking times for the two types of test trials were not significantly different.

Overall, Experiments 3 – 6 examined rule learning and generalization in infants from the training input containing some noise. It was found that infants generalized the rules to novel examples when rule-conforming sentences in the training had a higher type frequency than noise sentences (Experiments 3 and 5). When the type frequency was equal for rule and noise sentences in the training, infants' generalization was impeded (Experiments 4 and 6). We observed the same effect of type frequency with (in Experiments 3 and 4) and without morphological markings (in Experiments 5 and 6).

2.3 Rule generalization and the role of morphological markings when the level of noise is high and the exposure to the input is increased

2.3.1 Experiment 7: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; morphologically marked; increased training

In Experiment 7 we asked whether the learning and generalization would be successful in when rule types are not more frequent than noise types (as in Experiment 4) if the overall exposure to the same training set was increased.

2.3.1.1 Participants and Materials

Sixteen infants aged 14 months from various linguistic backgrounds completed the experiment. The age of the 9 boys and 7 girls ranged from 14 months 1 day to 15 months 1 day (*M* = 14 months 20 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Seven other infants were tested but their data not included in the analysis for various reasons such as fussiness (2), crying (1), parental interference (2), and looking to the screen for 2 seconds and less during 6 or more test trials (2). One other infant did not complete the experiment.

Materials were identical to those in Experiment 4.

2.3.1.2 Design and Procedure

Design and Procedure were identical to Experiment 4. The only difference was that the training stimuli were played twice (this training will hereafter be called 'double training'). There was a brief pause before the training stimuli were played a second time.

2.3.1.3 Results

Each infant's looking times during the two test trial types (trained rule versus non-trained rule) were calculated. A Paired Samples *t*-test revealed that infants showed significant discrimination between these two rules, $t(15) = -2.56$, $p = 0.022$, *two-tailed, partial eta squared* = 0.304. Cumulative looking time across test trials was on average 38.5 s (*SE* = 4.89) for the trained movement rule and 44.63 s (*SE* = 5.16) for the non-trained movement rule. Average looking time per trial was 7.7 s (*SE* = 0.98) for the trained movement rule and 8.93 s (*SE* = 1.03) for the non-trained movement rule.

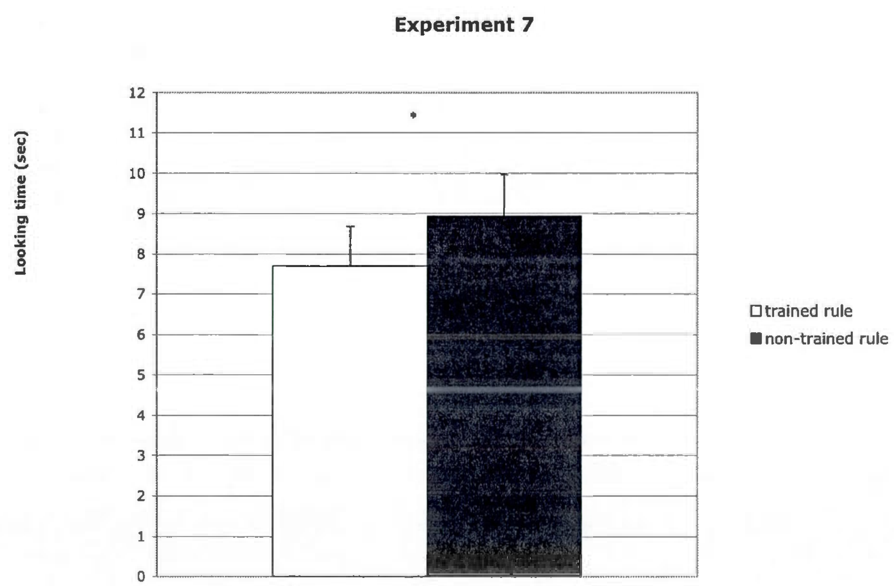

Figure 2.7 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 7: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of

noise 1:4; test – novel instances; morphologically marked; increased training. Infants' looking times for the two types of test trials were significantly different.

The results of Experiment 7 suggest that infants learned the rules. However, there is another possibility. Infants may have just paid attention to the movement of specific morphological endings because of increased overall exposure of the training set, without learning the rules that applied to novel stems. Experiment 8 hence tested whether the same input distribution would lead to rule learning in the absence of morphological markings.

2.3.2 Experiment 8: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; no morphological markings; increased training

In Experiment 8 we examined whether increased exposure to the training input with a non-dominant type frequency for rule instances could also lead to rule learning and generalization in the absence of morphological markings. The distributional properties of the input were the same as in Experiment 6. Only the exposure to the training stimuli was doubled.

2.3.2.1 Participants and Materials

Sixteen infants from various linguistic backgrounds completed the experiment. The age of the 8 boys and 8 girls ranged from 14 months 10 days to 15 months 6 days (*M* = 14 months 24 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Seven other infants were tested but their data not included in the analysis for various reasons such as fussiness (3), getting out of camera field during test trials (1), parental interference (2), and looking to the screen for 2 seconds and less during 6 or more test trials (1). One other infant did not complete the experiment.

Training and test stimuli were identical to Experiment 6. The only difference was that the training set was presented twice, with a brief pause before the stimuli were played for the second time.

2.3.2.2 Design and Procedure

Design and procedure were identical to those in Experiment 7.

2.3.2.3 Results

Each infant's looking times during the two test trial types (trained rule versus non-trained rule) were calculated. A Paired Samples *t*-test revealed that infants did not distinguish between these two rules, $t(15) = -1.07$, $p = 0.3$, *two-tailed, partial eta squared* = 0.071. Cumulative looking time across test trials was on average 40.11 s (*SE* = 5.63) for the trained movement rule and 44.29 s (*SE* = 5.44) for the non-trained movement rule. Average looking time per trial was 8.02 s (*SE* = 1.13) for the trained movement rule and 8.86 s (*SE* = 1.09) for the non-trained movement rule.

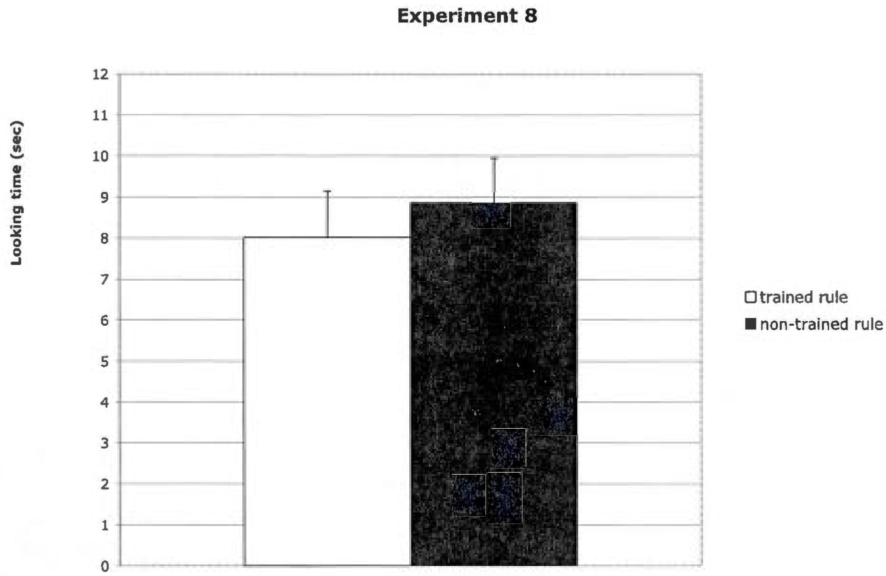

Figure 2.8 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 8: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; no morphological markings; increased training. Infants' looking times for the two types of test trials were not significantly different..

The two experiments with doubled training and non-dominant type frequency of rule exemplars in the training showed different results: infants in Experiment 7 discriminated between the trained and non-trained rules, whereas infants in Experiment 8 did not. The crucial difference between the stimuli in these two experiments was morphological makings: they were present in Experiment 7, but not in Experiment 8. What exactly did infants learn in Experiment 7? Did they just track the relations between specific morphological markings, without tracking the abstract relations of the whole words? Or, did they track the positional relations of the whole words, and morphological markings assisted this learning? In order to answer these questions, we designed Experiment 9. Here, infants were trained with morphologically marked input, and tested with unmarked stimuli.

2.3.3 Experiment 9: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; training morphologically marked, test unmarked; increased training

2.3.3.1 Participants and Materials

Sixteen infants from various linguistic backgrounds completed the experiment. The age of the 6 boys and 10 girls ranged from 14 months 5 days to 15 months 5 days (*M* = 14 months 26 days). Parents were asked about their children's language background (see p, 161, Appendix N). None of the infants had any prior exposure to Russian. Eight other infants were tested but their data not included in the analysis for various reasons such as fussiness (4), parental interference (2), experimenter's error (1), and looking to the screen for 2 seconds and less during 6 or more test trials (1). One other infant did not complete the experiment.

Training stimuli were identical to Experiment 7; test stimuli were identical to those from Experiment 8.

2.3.3.2 Design and Procedure

Design and procedure were identical to Experiments 7 and 8.

2.3.3.3 Results

Each infant's looking times during the two test trial types (trained rule versus non-trained rule) were calculated. A Paired Samples *t*-test revealed that infants did not discriminate the two rules, $t(15) = 1.16$, $p = 0.265$, *two-tailed, partial eta squared* = 0.082. Cumulative looking time across test trials was on average 35.19 s (*SE* =

4.67) for the trained movement rule and 30.33 s (*SE* = 3.93) for the non-trained movement rule. Average looking time per trial was 7.04 s (*SE* = 0.93) for the trained movement rule and 6.07 s (*SE* = 0.79) for the non-trained movement rule.

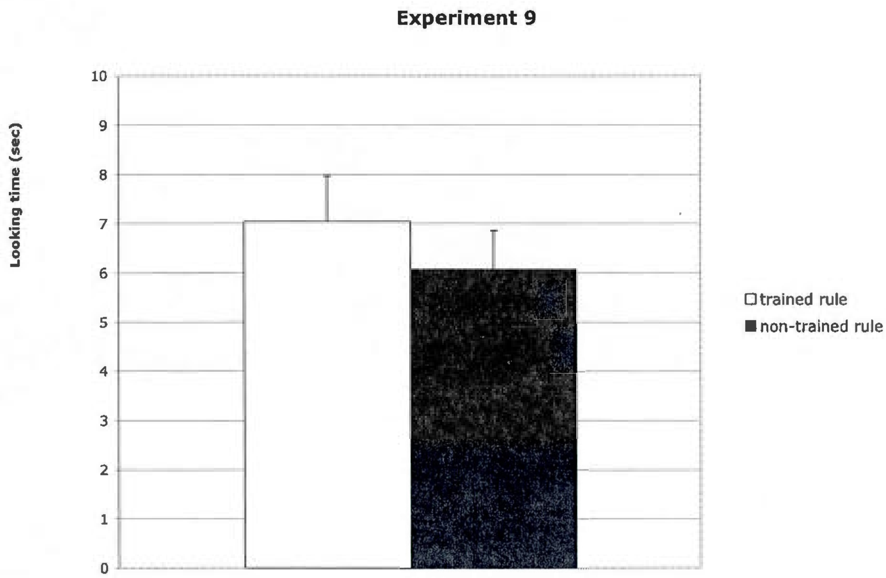

Figure 2.9 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 9: Training – 50% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – novel instances; training morphologically marked, test unmarked; increased training. Infants' looking times for the two types of test trials were not significantly different.

The null results of Experiment 9 excluded the interpretation that infants were tracking the abstract movement relations of the whole words. If that interpretation were correct, the results of Experiment 9 would have been positive. It is certain that positive results in Experiment 7 showed infants' learning of morphological markings. It is not clear, however, whether they just tracked specific morphological endings and their alternations without deriving anything abstract, or they derived an abstract rule that could apply to novel roots but required the markings.

In Experiment 10, we trained and tested infants with the same stimuli as in Experiment 9, except that we removed morphological markings in a subset of rule sentence pairs during training. This unmarked subset of training sentences was designed to show infants that morphological markings were not an obligatory part of the rule.

2.3.4 Experiment 10: Training – 50% types of rule instances; type-token ratio of rule sentences 1:4, type-token ratio of noise 1:4; test – novel instances; training partially marked, test unmarked; increased training

2.3.4.1 Participants and Materials

Sixteen infants (5 boys and 11 girls) from various linguistic backgrounds completed the experiment. The age ranged from 14 months 9 days to 15 months 13 days (M = 14 months 28 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Eight other infants were tested but their data not included in the analysis for various reasons such as fussiness (4), getting out of camera field during test trials (2), lack of interest (1) and looking to the screen for 2 seconds and less during 6 or more test trials (1). Four other infants did not complete the experiment.

Training stimuli (see p. 153 and p. 158, Appendices F and K) were identical to Experiment 7 except for two rule sentences taken from Experiment 8. Six rule pairs from Experiment 7 had morphological markings, i.e. all words in the original ABC sentences were marked by –a for the A position, –it for the B position, and –ku for the C position. Two rule pairs taken from the training of Experiment 8 did not have those markings.

The number of recordings and the arrangement of the training strings were the same as in Experiment 4 (and Experiment 7).

In the training, average sentence duration was 2.47 s (*SD* = 0.2) for Rule 1, 2.49 s (*SD* = 0.2) for Rule 2, and 2.39 s (*SD* = 0.15) for noise sentences. In the test, average sentence duration was 2.55 s (*SD* = 0.09) for Rule 1 and 2.55 s (*SD* = 0.09) for Rule 2. As in the previous experiments, the original and the moved version of a sentence within a pair were separated by approximately 700 ms. The pause between rule and noise types, between any pairs, and between any two noise sentences was approximately 1200 msec.

Test stimuli were identical to Experiment 8.

2.3.4.2 Design and Procedure

Design and Procedure were identical to Experiments 7, 8 and 9. The total duration of one training phase was 336.8 s for the 'Rule 1 + Noise' training condition and 334.4 s for the 'Rule 2 + Noise' training condition. Each training set was presented twice, with a brief pause between.

2.3.4.3 Results

Each infant's looking times during the two test trial types (trained rule versus non-trained rule) were calculated. A Paired Samples t-test found no significant difference between these two rules, $t(15) = 0.42$, $p = 0.681$, *two-tailed, partial eta squared* = 0.012. Cumulative looking time across test trials was on average 40.79 s (*SE* = 4.96) for the trained movement rule and 39.54 s (*SE* = 5.37) for the non-trained movement rule. Average looking time per trial was 8.16 s (*SE* = 0.99) for the trained movement rule and 7.91 s (*SE* = 1.07) for the non-trained movement rule.

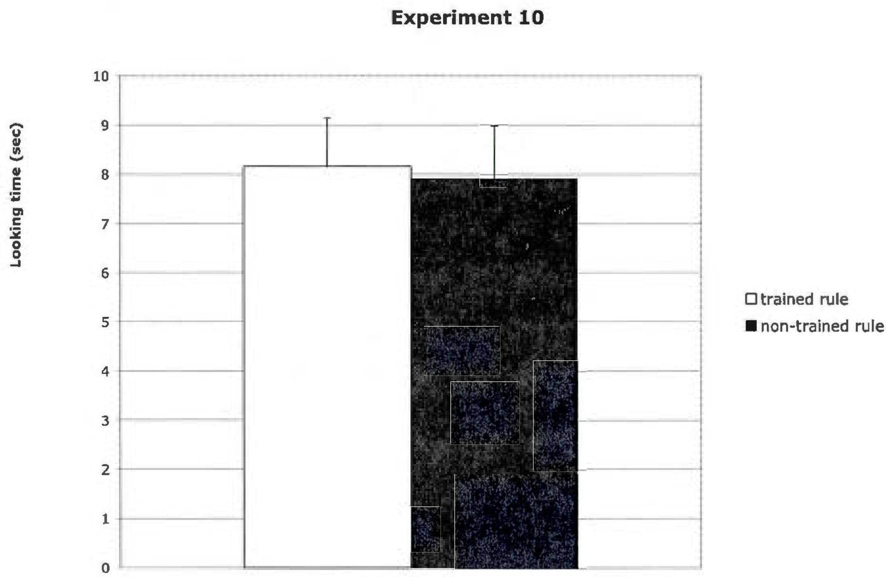

Figure 2.10 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 10: Training – 50% types of rule instances; type-token ratio of rule sentences 1:4, type-token ratio of noise 1:4; test – novel instances; training partially marked, test unmarked; increased training. Infants' looking times for the two types of test trials were not significantly different.

In Experiment 10, infants were trained with the input containing a subset of unmarked rule instances..These served as evidence that morphological markings were not obligatory for the abstract word-position movement rule. Infants did not show learning.

In Experiments 7 – 10 the frequency of rule types did not differ from the frequency of noise types. The results suggest that increased exposure (double training) of this kind of input combined with morphological markings led to learning. Although infants in Experiment 7 might have learned abstract word-position movement rules that required markings as part of the rule (i.e., abstract at the root with specific markings), it is more likely that they simply tracked specific morpheme relations without deriving anything abstract about the root. We thus had no clear evidence that infants were able to learn the abstract movement rules in the low type frequency situation.

2.4 The nature of rule encoding

2.4.1 Experiment 11: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; training morphologically marked, test unmarked

In Experiment 11 we examined the nature of infants' rule encoding and generalization from morphologically marked input. The combined results of Experiments 3-6 showed that infants can generalize rules in an abstract way when the input contained high rule type frequency. In Experiment 3, however, the test sentences shared the same morphological markings as the training input. It is not clear whether the results of that experiment show a broader generalization, or a more narrow generalization. In the broader generalization, any novel words are subject to the rule, whether they have the same morphological markings or not. In the narrower generalization, only words with the appropriate morphological markings should be subject to the rule. That is, the rule can apply to new words with novel roots but the words must have the morphological markings of the training input. In Experiment 11 we tested the broader generalization by training infants with morphologically marked input and testing them with unmarked sentences.

The distributional properties of the input were the same as in Experiment 3. Training stimuli were those of Experiment 3, eight rule pairs and two noise sentences. Therefore, the type frequency was 80% for rule instances and 20% for noise. Each rule pair occurred four times, i.e. the type-token ratio was 1:4. Each noise sentence occurred 16 times, i.e., the type-token ratio was 1:16. The training words were all morphologically marked. Test stimuli were taken from Experiment 5 (without markings).

2.4.1.1 Participants and Materials

Sixteen infants (5 boys and 11 girls) from various linguistic backgrounds completed the experiment. The age ranged from 14 months 11 days to 15 months 19 days (M = 14 months 31 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Four other infants were tested but their data not included in the analysis for various reasons such as fussiness (1), lack of interest (1), experimenter's error (1), and looking to the screen for 2 seconds and less during 6 or more test trials (1). One other infant did not complete the experiment.

Training stimuli were identical to training stimuli in Experiment 3; testing stimuli were identical to the testing stimuli in Experiment 5.

2.4.1.2 Design and Procedure

Design and procedure were identical to those ones in Experiments 3 – 6.

2.4.1.3 Results

Each infant's looking times during the two test trial types (trained rule versus non-trained rule) were calculated. A Paired Samples *t*-test revealed that infants did not show significant discrimination between these two rules, $t(15) = -0.8$, $p = 0.436$, *two-tailed, partial eta squared* = 0.041. Cumulative looking time across test trials was on average 41.91 s (*SE* = 5.03) for the trained movement rule and 45.44 s (*SE* = 5.22) for the non-trained movement rule. Average looking time per trial was 8.38 s (*SE* = 1.01) for the trained movement rule and 9.09 s (*SE* = 1.04) for the non-trained movement rule.

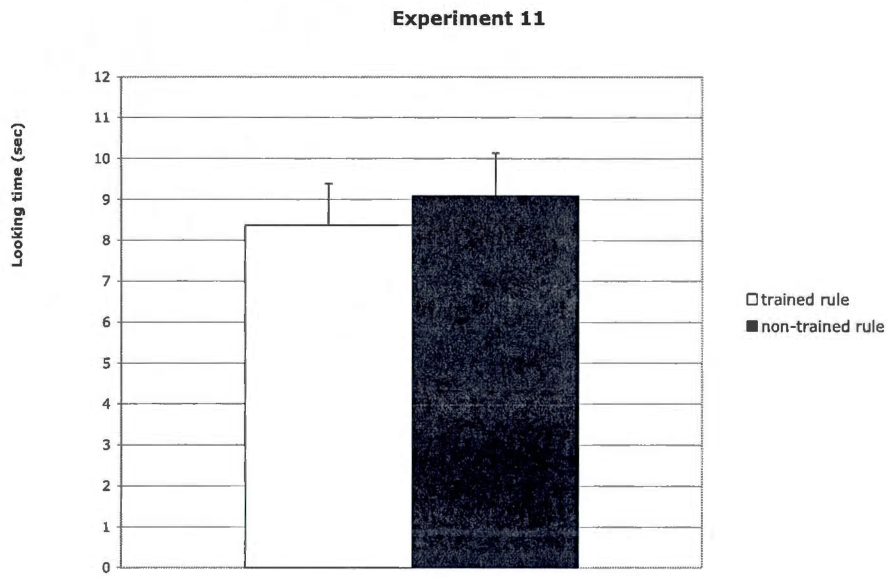

Figure 2.11 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 11: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; training morphologically marked, test unmarked. Infants' looking times for the two types of test trials were not significantly different.

These null results suggest that the broader generalization did not take place. Given the morphologically marked training input, infants did not apply the learned rule to novel words without markings. This suggests that when exposed to input with morphological markings, infants made a narrow generalization: only sentences with words sharing the same markings should be subject to the rule. Experiment 12 tested whether infants would make the broader generalization if the training input included a subset of unmarked rule sentences.

2.4.2 Experiment 12: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; training partially marked, test unmarked

In Experiment 12 we examined whether infants would make a broader generalization if not all rule instances in the training were morphologically marked. We first trained infants with partially marked examples. They were then tested with unmarked stimuli. If such training leads to broader generalization, infants should discriminate between the trained and non-trained rules in the test.

As in Experiment 11, training in Experiment 12 consisted of eight rule pairs (see p. 158, Appendix K) and two noise sentences (see p. 152, Appendix E). Type frequency was hence 80% for rule sentences and 20% for noise. Each rule pair occurred four times, i.e. type-token ratio was 1:4. Each noise sentence occurred 16 times, i.e., the type-token ratio was 1:16. Six out of eight rule pairs in the training had morphological markings (from among the training stimuli in Experiment 3). Two other rule pairs were unmarked (from the training stimuli in Experiment 5). Test stimuli were the same as the test stimuli in Experiment 5.

2.4.2.1 Participants and Materials

Sixteen infants from various linguistic backgrounds completed the experiment. The age of the 8 boys and 8 girls ranged from 14 months 7 days to 15 months 16 days (*M* = 14 months 22 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Eleven other infants were tested but their data were not included in the analysis for various reasons such as fussiness (5), crying (1), getting out of camera field during test trials (1), parental interference (3) and experimenter error (1). Two other infants did not complete the experiment.

Training stimuli were identical to Experiment 3 except for 2 rule pairs taken from Experiment 5. As in Experiment 3, each training set consisted of four strings. The number of recordings and the arrangement of the training strings were the same as in Experiment 3.

Within a sentence pair the original and the moved versions were separated by approximately 700 ms. The pause between rule and noise types, between any pairs, and between any two noise sentences was approximately 1200 ms.

Test stimuli were identical to those ones in Experiment 5.

In the training, average sentence duration was 2.48 s (*SD* = 0.2) for Rule 1 and 2.49 s (*SD* = 0.2) for Rule 2.

2.4.2.2 Design and Procedure

Design and procedure were identical to Experiments 3 – 6. Total duration of the training phase was 337 s for 'Rule 1 + Noise' training and 336.6 s for 'Rule 2 + Noise'.

2.4.2.3 Results

Each infant's looking times during the two test trial types (trained versus non-trained rules) were calculated. A Paired Samples *t*-test revealed that infants showed significant discrimination between these two rules, $t(15) = -2.48$, $p = 0.026$, *two-tailed, partial eta squared* = 0.29. Cumulative looking time across test trials was on average 35.44 s (*SE* = 4.61) for the trained movement rule and 41.31 s (*SE* = 4.91) for the non-trained movement rule. Average looking time per trial was 7.09 s (*SE* = 0.92) for the trained movement rule and 8.26 s (*SE* = 0.98) for the non-trained movement rule.

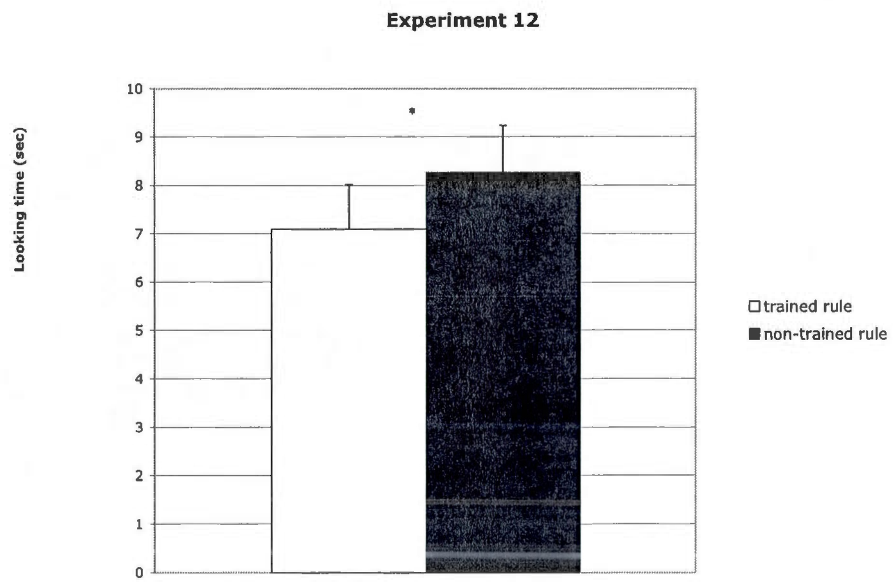

Figure 2.12 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 12: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – novel instances; training partially marked, test unmarked. Infants' looking times for the two types of test trials were significantly different.

These positive results suggest that given input with optional morphological markings on the rule stimuli, infants encoded a broader movement rule which did not require markings in novel instances.

2.5 Over-generalization and learning of exceptions

2.5.1 Experiment 13: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – noise instances from the training; morphological markings

In the previous series of experiments, we examined different factors in infants' learning of movement rules and their generalization to novel instances in the test. While examining the role of relative type frequency of rule instances for rule abstraction, we introduced the noise sentences and manipulated their variability across the experiments. The noise consisted of examples that did not apply any rule. Thus, they were not overt violations. This kind of noise has particular interest. It can in principle be interpreted as true noise, i.e. cases to which the rule should never apply. It can also be interpreted as cases which did not yet have a chance to be heard with the rule, but applying the rule to them would be grammatical. The above experiments suggested that even the cases of non-applied rules were interpreted as true noise.

In Experiments 13 – 15 we examined the conditions under which infants would interpret noise as exceptions or as cases subject to the rule, i.e., over-generalization.

To examine these two processes (i.e., memorizing exceptions versus rule over-generalization), we designed experiments in which the trained and non-trained rules were applied in the test to the noise sentences from the training. If infants can discriminate the rules applied to such noise sentences from training, it means that they apply the trained rule to them, i.e., they over-generalizate. If infants instead fail to distinguish the noise sentences in which the the two rules are applied, it means that they treat the noise as exceptions.

In Experiment 13, we used an input condition which had already shown rule generalization to novel instances (Experiment 3). The training consisted of eight rule pairs and two noise sentences. Type frequency was hence 80% for rules and 20% for noise. Each rule pair occurred four times, and each noise sentence occurred 16 times. That is, the token frequency per type for noise was much higher than for ruleful examples.

In the test, the trained and non-trained rules were applied to the noise sentences from the training, rather than to novel sentences. In the training, those noise sentences were highly repetitive, i.e., they had a high token frequency per type. We predicted that high token frequency per type would guide infants to treat those noise sentences as exceptions, i.e. cases to which the rule should not apply. That is, noise with a high type-token ratio from the training should resist over-generalization.

2.5.1.1 Participants, Materials, Design and Procedure

Sixteen infants from various linguistic backgrounds completed the experiment. The age of the 10 boys and 6 girls ranged from 14 months 1 day to 14 months 23 days (*M* = 14 months 16 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Five further infants were tested but their data were not included for various reasons such as fussiness (1), crying (1), experimenter error (1), and looking to the screen for 2 seconds and less during 6 or more test trials (2). Three other infants did not complete the experiment.

Training materials were taken from Experiment 3. Test stimuli were the noise sentences from the training set (see p. 159, Appendix L). While in the training those noise sentences did not go through any movement, in the test phase the trained versus non-trained rules were applied to them. Each noise sentence was recorded with both kinds of movement. The original ABC noise sentences used in the test trials were new recordings which were different from those in the training. One recording was used for each test sentence and for each of its inverted versions.

For counterbalancing purposes in the test phase, one group of infants heard Rule 1 with one noise sentence and Rule 2 with another noise sentence, the other group group heard the reverse. The two trial types alternated during the test. The order of the two trial types was counterbalanced across infants: one group heard the trained rule first, the other the untrained one. These counterbalancing characteristics were comparable to Experiment 3. As in the previous experiments, the ISI within sentence pairs was approximately 700 ms, whereas between pairs it was approximately 1200 ms. The average duration of test sentences was 2.44 s (SD = 0.01) for Rule 1 and 2.47 s (SD = 0.04) for Rule 2. All other aspects of the design and procedure were identical to Experiment 3.

2.5.1.2 Results

Each infant's looking times during the two test trial types (trained versus non-trained rules) were calculated. A Paired Samples t-test revealed no discrimination between the two rules, $t(15) = -0.19$, $p = 0.854$, *two-tailed, partial eta squared* = 0.002. Cumulative looking time across test trials was on average 36.29 s (SE = 4.3) for the trained rule and 37.17 s (SE = 4.86) for the non-trained one. Average looking time per trial was 7.26 s (SE = 0.86) for the trained rule and 7.43 s (SE = 0.97) for the non-trained.

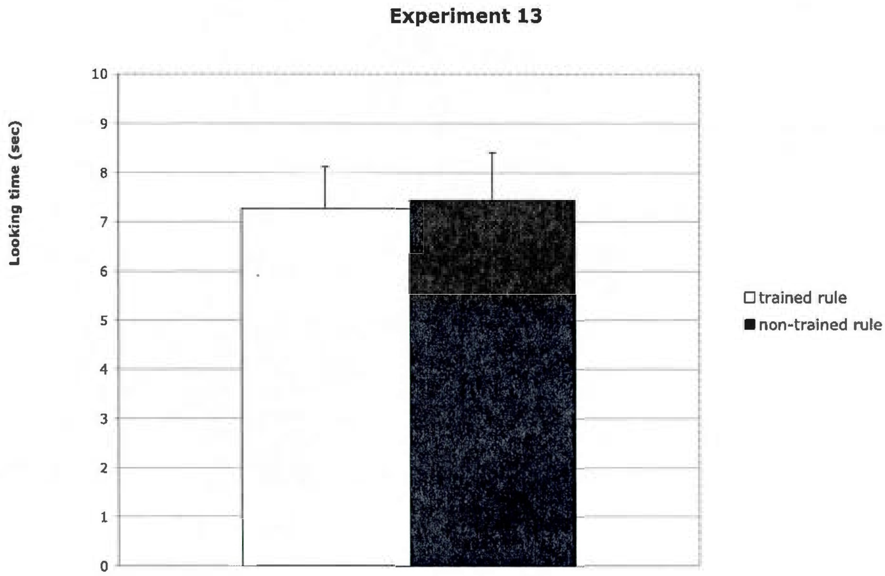

Figure 2.13 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 13: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:16; test – noise instances from the training; morphologically marked. Infants' looking times for the two types of test trials were not significantly different.

These results show that under the condition favorable for rule generalization to novel instances (as shown in Experiment 3), infants in Experiment 13 did not apply the rule they had acquired to noise examples that were highly repetitive (i.e., high token frequency per type) in the training; thus, they resisted over-generalization. This suggests that high noise token frequency led to the learning of those cases as exceptions.

The combined results of Experiments 3 and 13 confirmed our prediction that high rule type frequency of rule (relatively to noise type frequency), led to rule abstraction, whereas high token frequency per type for noise led to the learning of exceptions.

We then examined the conditions under which noise would not be learned as exceptions, but rather, as being eligible for applying the rule, i.e., over-generalization. We suggest that when type-token ratio is low for noise instances, they tend to be over-generalized to the learned rule. Experiment 14 tested this.

2.5.2 Experiment 14: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – noise instances from the training; morphologically marked

In Experiment 14, infants were trained with rule sentences that were 80% by type and by overall frequency. The training consisted of eight rule sentence pairs and two noise sentences. Type frequency was hence 80% for rule pairs and 20% for noise. Each rule pair occurred four times (type-token ratio 1:4). Crucially, each noise sentence occurred four times (type-token ratio 1:4), instead of 16 times in Experiments 3 and 13. In the test phase, the trained and the non-trained rules were applied to noise sentences from the training.

2.5.2.1 Participants, Materials, Design and Procedure

Sixteen infants from various linguistic backgrounds completed the experiment. The age of the 9 boys and 7 girls ranged from 14 months 1 day to 15 months 2 days (*M* = 14 months 14 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Fifteen other infants were tested but their data were not included in the analysis for various reasons such as fussiness (3), getting out of camera range during test trials (1), crying (1), toy distraction during test trials (1), lack of interest (1), parental interference (2), experimenter error (4), and looking at the screen for 2 seconds or less during 6 or more test trials (2).

Rule instances and their distribution across the training strings were identical to Experiment 13 (and 3). The same two noise sentences from Experiment 13 (and 3) were used in the training of Experiment 14. Two recordings of each noise sentence were taken from the multiple recordings of noise in Experiment 13 (and 3). One of the examples was used in Strings 1 & 3; the other in 2 & 4. Overall, each string contained one occurrence of the eight rule pairs and one occurrence of the two noise sentences. In total, the four strings for each of the two training sets consisted of eight rule pairs plus the two noise instances four times each. Test stimuli were identical to those in Experiment 13.

In the training, average sentence duration was 2.49 s (SD = 0.2) for Rule 1, 2.51 s (SD = 0.21) for Rule 2, and 2.37 s (SD = 0.08) for noise sentences.

Procedure and test phase design were identical to Experiment 13. Total duration of the training phase was 254 s for both 'Rule 1 + Noise' and 'Rule 2 + Noise' training conditions.

2.5.2.2 Results

Each infant's looking times during the two test trial types (trained versus non-trained rules) were calculated. A Paired Samples t-test revealed that infants discriminated these two rules, $t(15) = -3.07$, $p = 0.008$, *two-tailed, partial eta squared* = 0.385. Cumulative looking time across test trials averaged 35.66 s (SE = 3.78) for the trained movement rule and 44.47 s (SE = 4.45) for the non-trained movement rule. Average looking time per trial was 7.13 s (*SE* = 0.76) for the trained movement rule and 8.89 s (*SE* = 0.89) for the non-trained.

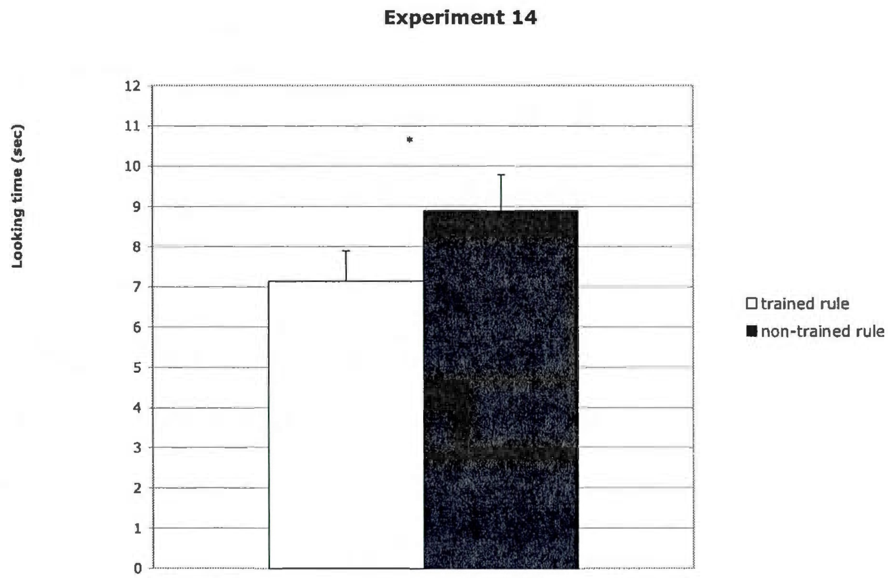

Figure 2.14 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 14: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – noise instances from the training; morphologically marked. Infants' looking times for the two types of test trials were significantly different.

These results demonstrate over-generalization, i.e. the application of the rule to the exceptions. The results contrast with those of Experiment 13, in which infants did not show discrimination. The difference across these two experiments was the type-token ratio of noise instances in the training. We can conclude that the factor which led infants to over-generalization was the low token frequency of noise instances in the training. When the token frequency of noise instances in the training was high, infants memorized the exceptions (Experiment 13).

Stimuli in Experiments 13 and 14 were morphologically marked in both training and test. It is possible that morphological markings assisted infants' overgeneralization in Experiement 14. In order to test this possibility, we designed Experiment 15, identical to the design of Experiment 14 in all respoects, except that it did not have morphological markings.

2.5.3 Experiment 15: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – noise instances from the training; no morphological markings

In Experiment 15 we asked whether over-generalization of syntactic movement can be affected by the absence of morphological markings.

2.5.3.1 Participants and Materials

Sixteen infants aged 14 months from various linguistic backgrounds completed the experiment. The age of the 7 boys and 9 girls ranged from 13 months 30 days to 15 months 13 days (M = 14 months 24 days). Parents were asked about their children's language background (see p. 161, Appendix N). None of the infants had any prior exposure to Russian. Six other infants were tested but their data were not included in the analysis for various reasons such as fussiness (3), crying (2) and sleepiness (1).

Rule instances and their arrangement across the training strings were identical to Experiment 5. The same two noise sentences from Experiment 5 were used. Two recordings of each noise sentence were chosen; each exemplar occurred twice across the training strings. The arrangement of rule and noise instances in the training strings was the same as in Experiment 14. In total, the four strings contained the eight rule pairs and the two noise instances four times each. Rule instances were most frequent by type (80%) relative to noise (20%).

As in Experiments 13 and 14, test stimuli were noise sentences from the training (that did not go through any movement), and they were now applied to both rules in the test trials (see p. 160, Appendix M). For the original ABC sentences the same speaker recorded new exemplars along with the inverted versions. One recorded example was used for each test sentence and for each of its inverted versions.

As in the previous experiments, the ISI within sentence pairs was approximately 700 ms and between pairs approximately 1200 ms. Average duration of noise sentences in the training was 2.67 s (SD = 0.19). Average duration of test sentences was 2.61 s (SD = 0.02) for Rule 1 and 2.6 s (SD = 0.01) for Rule 2. All other aspects of the design and procedure were identical to Experiment 14.

2.5.3.2 Procedure and Design

Procedure and design were identical to Experiment 14. Total duration of the training phase was 260 s for both 'Rule 1 + Noise' and 'Rule 2 + Noise' training.

2.5.3.3 Results

Each infant's looking times during the two test trial types (trained versus non-trained rules) were calculated. A Paired Samples t-test showed no discrimination beween the two rules, $t(15) = -0.27$, $p = 0.789$, *two-tailed, partial eta squared* = 0.005. Cumulative looking time across test trials averaged 44.27 s (SE = 5.71) for the trained movement rule and 45.83 s (SE = 4.35) for the non-trained movement rule.

Average looking time per trial was 8.85 s (*SE* = 1.14) for the trained movement rule and 9.17 s (*SE* = 0.87) for the non-trained one.

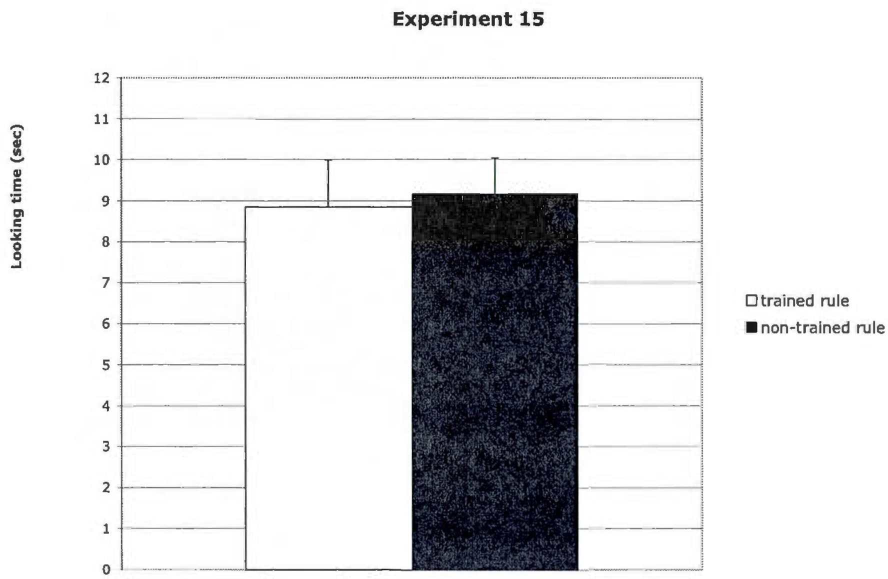

Figure 2.15 Mean and standard error of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 15: Training – 80% types of rule instances; type-token ratio of rule instances 1:4, type-token ratio of noise 1:4; test – noise instances from the training; no morphological markings. Infants' looking times for the two types of test trials were not significantly different.

In Experiment 15, we did not observe over-generalization, unlike the over-generalization found in Experiment 14. These results suggest that morphological markings play an important role in over-generalization. Even the low frequency of repetition of the noise in Experiment 15 (type-token ratio of 1:4) might have been too high and led to the learning of exceptions. In Experiment 14, in which the stimuli were morphologically marked, those markings could have led infants to accept over-generalizations more easily. Further research is needed to see whether fewer repetitions would lead infants to over-generalize in the absence of morphological markings.

DISCUSSION AND CONCLUSION

The experiments in this thesis addressed the question of how infants learn abstract rules and generalize them to novel instances. In particular, we examined the role of type and token frequencies in rule generalization, over-generalization and the learning of exceptions. In some experiments the goal was to understand what distributional properties of the input make infants learn abstract rules and generalize the trained rules to novel instances. We also examined the nature of the rule abstracted by infants from the input where two rules were possible. In other experiments we evaluated the conditions under which infants over-generalize the learned rule to examples of noise and the conditions under which they resist over-generalization and learn exceptions.

To test these questions, we designed two artificial word-order movement rules using Russian, a language unknown to our infant participants. In Rule 1, sentences with ABC order moved immediately into BAC sentences. In Rule 2, the ABC sentences moved into ACB sentences. Infants were trained with a set of sentences following one of the rules. In some experiments the training contained additional sentences that did not conform to the rule. These instances served as noise. In Experiments 1 – 12, we examined the generalization of learned rules to novel instances. Hence, test trials were novel instances conforming to the trained and untrained rules, respectively. In Experiments 13 – 15, we tested over-generalization and the learning of exceptions. Here, test trials were noise sentences from the training that did not go through any movement but were now applied to the trained rule versus untrained rule.

In Experiments 1 and 2, we aimed at exploring infants' capacity to learn and generalize abstract rules with ideal learning input, i.e., when 100% of input instances conformed to the rule. 11-month-olds in Experiment 1 were unable to learn. 14-month-olds in Experiment 2 learned the abstract rule and generalized it to novel instances.

The results of Experiment 2 are consistent with the findings by Marcus and colleagues (1999). Their study demonstrated that infants can learn abstract rules based on simple identity-based patterns at seven months. In our study, learning was successful with older infants. This could be due to the greater complexity of the stimuli. Our stimuli were designed with multisyllabic words, whereas Marcus and colleagues (1999) used monosyllabic words. Another reason could be the complexity of the rule. The rule in their study was a simpler one, based on identity patterns (e.g., AAB, ABA). Our rule involved the movement of words according to their positional categories. For Rule 1, infants had to learn that the first two positional categories in a sentence had to exchange their positions (e.g., ABC – BAC). They had to track each sentence in the original word order and the re-ordered version of that sentence. These properties could explain why generalization was successful only in 14-month-olds and not in 11-month-olds.

Experiments 3 – 6 examined the role of type frequency in infants' learning and generalization of movement rules. Here, the training input contained some noise. We designed experiments in which rule instances were dominant by type frequency over noise (Experiments 3 and 5), as well as experiments in which rule and noise instances were equal in type frequency (Experiments 4 and 6). In Experiments 3 and 4, all A, B and C words in both the training and test conditions were morphologically marked. In Experiments 5 and 6, the words were unmarked.

In both experiments in which the rule types were more frequent (Experiments 3 and 5), the infants learned, whereas in both experiments with equal type frequency (Experiments 4 and 6), they did not. These results suggest that infants can learn and generalize movement rules when input contains a certain level of noise, as long as the type frequency of the rule instances is higher than the noise. The absence of morphological markings in Experiments 5 and 6 did not affect infants' performance. Infants in Experiments 5 and 6 showed the same pattern of results as in Experiments 3 and 4 (which had morphological markings).

In Experiments 7 – 10 we further examined the possibility that morphological markings might help infants learn and generalize movement rules when the level of noise is high. Taken together, Experiments 7 – 10 suggest that under conditions of high noise input with repeated exposure to the same training input led infants learn specific morphological markings and their movement patterns (e.g., morph1 morph2 morph3 – morph2 morph1 morph3). There was no definitive evidence that infants learned the abstract rule learning and generalized it to novel stems with those morphological endings (i.e., A_{morph1} B_{morph2} C_{morph3} – B_{morph2} A_{morph1} C_{morph3}).

In general, Experiments 3 – 10 showed how generalization was affected by the proportional differences in type and token frequencies of rule and noise exemplars in the training. We found that infants' generalization was successful when rule-conforming instances were high in type and low in token frequencies relative to low-type high-token noise. When rule instances were no longer more frequent than noise, generalization was impeded. These results are consistent with a study by Gómez and LaKusta (2004). There, infants' generalization was also impeded when the input contained a lower number of types of rule instances. However, in their study, type frequency was not separated from the overall frequency. Across their experimental conditions, a dominant rule was reduced in both type and overall frequency. It was not clear from that study whether type frequency alone was important for rule abstraction. To control for overall frequency of rule and noise utterances, we designed experiments in such a way that overall frequency was always equal for rule and noise instances. Only type and token frequencies varied across experiments. We found that rule instances had to be high in type and low in token frequencies relative to noise instances in order to allow generalization.

In Experiments 11 and 12, we examined the nature of the rule encoded by infants in the high type, morphologically marked training condition (i.e., in Experiment 3). Experiment 3 suggested two possible kinds of rule generalizations: 1) a broader abstract movement rule that applied to any novel words (e.g., ABC – BAC); 2) a narrower abstract movement rule applicable to novel words, with the words requiring specific morphological markings (e.g., A_{morph1} B_{morph2} C_{morph3} – B_{morph2} A_{morph1} C_{morph3}). Recall that infants in Experiment 5 (which had the same input distribution as Experiment 3, but without morphological markings) learned a broader abstract rule. Experiment 11 thus examined whether the morphologically marked input in Experiment 3 would lead to the broader kind of abstract rules (e.g., ABC – BAC). Infants were trained with the same input as in Experiment 3 (with markings) and tested with unmarked test stimuli from Experiment 5. The results showed no evidence that infants learned the broader rule. The results of Experiments 3 and 11 suggest that when rule exemplars were morphologically marked, infants made the narrower generalization.

Experiment 12 examined whether a few unmarked rule instances in the training would be sufficient for infants to encode the broader movement rule. The training input was the same as in Experiments 11 and 3, except that a subset of rule instances had no morphological markings. The test stimuli were unmarked, as in Experiment 11. Infants in Experiment 12 demonstrated the learning of the broader rule. The results of Experiments 3, 5, 11 and 12 showed that rule generalization/abstraction encoded the morphological markings (if they occurred in the input) as well as the degree of consistency of their presence. Experiment 12 also contrasts with the null results of Experiment 10, in which the type frequency of noise exemplars was high.

When two rule generalizations were possible in our input, infants chose the more conservative one. These results are consistent with the study of Gerken (2006) where the training input allowed two generalizations, one broader and another more conservative. The broader generalization implied that the trained rule could be generalized to any novel instances for all parts of the rule. The more conservative generalization included a specific item in the input that could not be extended to other novel items. Infants made the more conservative generalization. The training input in the study of Gerken (2006) did not contain any noise examples. Our results suggest that even in the presence of noise, infants' generalization is more conservative.

In a subsequent study by Gerken (2010), a few counterexamples were added in the training input. With these counterexamples, infants made a broader generalization (e.g., AAB and not AAdi). These results are compatible with the results of Experiment 12. Here, when the training input contained a few morphologically unmarked cases, infants made a broader generalization. Unlike the study of Gerken (2010), stimuli in Experiment 12 also contained some noise sentences. It can be suggested that the presence of noise did not change the pattern of learning.

Experiments 13 – 15 examined infants' learning of exceptions. Experiments 3, 5 and 12 showed that type frequency was important for rule generalization. Noise instances, which were repeated often during training in these experiments, did not impede rule abstraction. In Experiment 13, we tested the possibility that those frequently repeated noise instances were learned by infants as exceptions to the movement rule. The training input was identical to that of Experiment 3. The trained and non-trained rules in the test were applied not to novel sentences, but to the highly repeated noise sentences from the training. Since Experiment 3 showed infants' ability to generalize abstract rules, we could assume that in Experiment 13 infants should also learn the abstract rule. The particular interest of Experiment 13 was the learning of exceptions versus over-generalization. Positive results in this experiment would demonstrate an over-generalization of the trained rule to those noise sentences. Null results would show no over-generalization, hence, the learning of exceptions.

Infants in Experiment 13 did not discriminate between the trained and non-trained rules that were applied to the noise sentences that they had heard during training. These results suggest that high type-token ratio (i.e., high token frequency) of noise instances blocked over-generalization and led to the learning of those instances as exceptions. Following the same idea, Experiment 14 tested whether the low token frequency of noise instances would lead to over-generalization. In Experiment 14 the token frequency of each noise sentence was reduced to 4, in contrast to 16 in Experiment 13. Infants in Experiment 14 discriminated the noise being applied to the trained rule versus to the untrained rule, which suggests that the low token frequency of noise led to the over-generalization of these instances to the rule.

Experiments 13 and 14 had morphological markings in both the training and the test phase. Experiment 15 examined whether the low token frequency of noise instances also led to over-generalization when the input contained no morphological markings. Experiment 15 had the same design as Experiment 14, except that there were no morphological markings in the training and the test phases. However, in Experiment 15 infants did not discriminate the test trials. Morphological markings in Experiment 14 seemed to have led infants to over-generalize noise instances more readily. When the learning input contained no markings as in Experiment 15, infants' attention may have been directed more toward specific details of noise instances, leading them to resist over-generalizing these instances. The token frequency of non-morphologically marked noise might have been high enough in Experiment 15 to trigger the learning of those cases as exceptions. Future research should test whether infants over-generalize when the token frequency of unmarked noise is further reduced.

The over-generalization observed in Experiment 14 is consistent with existing evidence of over-generalization in children's productions (e.g., Bowerman, 1988). Our work shows that infants can over-generalize perceptually before being able to speak. Moreover, we show that over-generalization can be distributionally driven. These results are consistent with the previous finding that children and adults make more over-generalization errors with low-frequency verbs (Brooks et al., 1999; Theakston, 2004; Ambridge et al., 2008).

The role of token frequency in the learning of exceptions is consistent with the findings of Wonnacott, Newport and Tanenhaus (2008). They demonstrated that adults, learning a miniature artificial language, entrenched high-frequency verbs to a verb argument structure in which they occurred in the training. They made more errors with low-frequency verbs. Similarly, studies with children and adults showed that high-frequency verbs are less likely to be subject to over-generalization errors, and are hence better entrenched in their specific occurrences in the input.

The role of high token frequency for learning specific items was shown indirectly by Gerken (2006). Infants were trained with three-word utterances that contained a final word which occurred frequently. Infants learned tracking that frequent final word. Other indirect evidence comes from a study on non-adjacent dependencies (Gómez, 2002). When a middle element occurred frequently, infants learned specific sequences.

Did infants in Experiment 3 simply track the movement patterns of specific morphological markings without learning the word-level abstract movement rule? Experiments 4 and 13 eliminated this interpretation. In Experiment 13, which focused on the learning of exceptions from the same input distribution as Experiment 3, the test stimuli contained the same markings as did the test stimuli in Experiment 3. The markings were also the same in the training of both experiments. Experiment 13 and Experiment 3 did not yield the same results. If infants were simply tracking specific patterns of morphological endings, they should have shown the same pattern of results in both Experiments 3 and 13. Similarly, Experiment 4 had the same frequency of morphological markings as Experiment 3 in both the training and testing stimuli. In fact, the tracking of the specific patterns of morphological endings could have been easier in Experiment 4, since the endings occurred with more variable roots for noise instances (than those in Experiment 3), thus providing better segmentation cues (i.e., transitional probability cues) for the endings in Experiment 4. However, infants did not discriminate the test trial stimuli in Experiment 4. Hence infants in Experiment 3 did not simply track morphological markings or their movement patterns. Rather, they learned the abstract movement rule and generalized it to novel words.

Overall, the series of experiments in this thesis shows the role of distributional factors in rule generalization and the learning of exceptions. In terms of rule learning and generalization, our results are consistent with the 'Discovery Procedures' Acquisition Model proposed by Braine (1971). According to this model, learners learn more frequent linguistic patterns first. However, this model did not make a distinction between different kinds of frequencies. Our results suggest that infants generalize an abstract rule to novel items when rule-conforming instances in the training are high in type frequency relative~~ly~~ to noise instances. When rule instances in the training are no longer higher in frequency, infants do not learn the abstract rule.

Distributional cues alone were sufficient for the learning and generalization of abstract rules. We used stimuli from a language unknown to our infant participants. Hence, for them, stimuli were deprived of any semantic and syntactic information. This suggests that the learning of abstract rules does not require semantic or syntactic information.

Braine's model allowed over-generalization errors at the early stages of learning. It assumed that children would apply the most general patterns to all cases. We found over-generalization in infants. Over-generalization was frequency based. It was driven by low token frequency of exceptions. When some items occurred infrequently in the input, infants over-generalized them to the dominant rule.

With respect to the learning of exceptions, our results are compatible with the Entrenchment Model, according to which words that occur frequently in a construction are "entrenched" to that construction (Braine & Brooks, 1995). In our study infants resisted applying the trained rule to non-application noise cases from the training when those noise cases occurred frequently in the training. This suggests that the learning of exceptions was based on high token frequency of noise exemplars. The Preemption Model (e.g., Clark, 1987; Markman, 1989; Pinker, 1984; Goldberg, 1995) does not apply to our results since we did not have a semantic component. Infants in our study are not yet at the full stage of learning word meaning, and our stimuli from the unknown language did not bear any semantic information for them. Our results therefore suggest that exceptions can be learned from distributions without requiring semantics.

The kind of noise we used was non-application of the rule, rather than overt rule-violations. Logically, these could be rule-conforming exemplars that had not yet been heard with the rule but that could possibly be heard in the future. However, infants' failure to learn and generalize the rule in experiments with many noise exemplars suggests that infants indeed treated the non-application cases as true noise. This suggests that infants are conservative in their learning.

On the other hand, infants' generalization was successful when the training contained few noise examples. One interpretation is that such low-frequency non-application cases were treated as true noise, although they did not affect generalization. Another interpretation is that those cases were not treated as true noise, due to their low type frequency.

Overall, we showed one cue that can lead to the solution to Baker's paradox, that is, distributional properties of the input. Non-application cases are learned as exceptions when their token frequency is high. Low token frequency of non-application cases leads to their over-generalization to the rule. This finding is consistent with the Entrenchment Model.

Our results are also partially compatible with the Discovery Procedures Acquisition Model proposed by Braine (1971). According to this model, frequency determines whether a pattern noticed by a learner in a linguistic input reaches the *'permanent memory store'* or stays in the *'intermediate store'* of the memory component. Only patterns occurring in multiple sentences reach the *'permanent memory store'*, thus providing learners with the immunity to unsystematic errors. Although our study does not address the question of what cognitive components are required for learning, the frequency effects are compatible with those predicted by the Discovery Procedures Acquisition Model. In particular, patterns with a high type frequency are learned. However, our findings do not agree with some statements of the Discovery Procedures Acquisition Model (Braine, 1971). In Braine's model, learning is sequential: first general patterns are learned, and only afterwards learners attain to specific examples, i.e., exceptions. Our results suggest that infants can simultaneously learn a general pattern and specific examples from the same training. More precisely, the training in Experiments 3 and 13 were identical. In Experiment 3, infants generalized the learned rule to novel examples. In Experiment 13, they resisted applying the rule to noise sentences from the training, thus suggesting the learning of exceptions. Our findings are compatible with other studies showing that adults and children can simultaneously track statistics of general patterns and specific instances (Wonnacott, Newport & Tenenbaum, 2008; Wonnacott, 2011).

One important aspect of this thesis is the use of morphologically marked stimuli. In some experiments, training and/or test stimuli were morphologically marked, whereas in other experiments, partially marked and/or unmarked. One possible expectation would be that morphological markings should assist infants' learning of word order movement rules, since markers are reliable features of positional categories of words in sentences. However, we did not find evidence that morphological markings assist learning of movement rules. Infants learned the rule from unmarked stimuli when the input did not contain any noise (in Experiment 2). Note that the stimuli in Experiment 2 contained more complex word forms than in Experiments 3 – 15. In particular, the words used in Experiment 2 had a variable number of syllables, sometimes more than two, and had no consistency in morphological markings. In the experiments with added noise, infants still learned the rule from both marked and unmarked stimuli (e.g., Experiments 3 and 5).

An alternative interpretation could be considered for our experiments. This interpretation is based on the general observation in natural languages that morphologically rich languages are more flexible in their word order. It is possible that when stimuli are unmarked, infants pay more attention to the words themselves and their movement, similarly to languages without markings. In particular, in Experiment 2 infants learned the word order movement rule despite the complexity of stimuli. This interpretation may be applied to the results of experiments on overgeneralization of exceptions (Experiments 14 and 15). In Experiment 14, where both training and test were morphologically marked, infants applied the rule to noise sentences from the training. In the experiment with unmarked stimuli (Experiment 15), they did not apply the rule to noise sentences from the training. It is possible that in Experiment 15 with unmarked stimuli, infants paid more attention to individual words and hence learned specific noise sentences despite a relatively low number of repetitions in the training. Thus, they resisted over-generalizing them to the trained rule. In Experiment 14 with marked stimuli, infants paid less attention to words and did not learn specific noise sentences well enough. This resulted in the over-generalization of those noise sentences to the trained rule.

To further examine how morphology affects learning of word order, we conducted an additional analysis. Given the general observation in natural languages that morphologically richer languages have freer word order, we can predict that learning of word order movement rules should be better in morphologically unmarked experiments. We tested this prediction with ANOVAs of Experiments 3 and 4, and Experiments 5 and 6. In Experiments 3 and 4, both training and test stimuli were morphologically marked. In Experiments 5 and 6, all stimuli were unmarked.

The first analysis was performed on Experiments 3 and 4, which had morphological markings. In Experiment 3, rule sentences in the training had a higher type frequency than noise sentences, whereas in Experiment 4 both rule and noise sentences in the training were equal by type frequency. A 2x2 Mixed-design ANOVA was conducted to compare the effect of the rule familiarity on looking times in Experiments 3 and 4. The dependent variable was infants' looking time during test trials. The within-factor was Test Rule Type (trained vs. non-trained), whereas the between-factor was Type Frequency of Rule Instances (80% in Experiment 3 and 50% in Experiment 4). No significant within-subject effect of Test Rule Type was found, $F(1, 30) = 2.506$, $p = 0.124$, *partial eta squared* = 0.077. The between-subject effect of Type Frequency was not significant, $F(1, 30) = 0.011$, $p = 0.919$, *partial eta squared* = 0. The interaction between Test Rule Type and Type Frequency had a tendency for significance, $F(1, 30) = 3.461$, $p = 0.073$, *partial eta squared* = 0.103.

The second analysis was performed on Experiments 5 and 6 where all stimuli were morphologically unmarked. They had the same number of types and tokens of rule and noise sentences in the training as in Experiments 3 and 4. In Experiment 5, rule sentences had a dominant type frequency in the learning input, whereas in Experiment 6, rule and noise instances in the training were equal by type frequency. Again, a 2x2 Mixed-design ANOVA was conducted to compare the effect of the rule familiarity on looking times across these two experiments. Like in the previous analyses, the dependent variable was infants' looking time during test trials. The within-factor was Test Rule Type (trained vs. non-trained), whereas the between-factor was the Type Frequency of Rule Instances (80% in Experiment 5 and 50% in Experiment 6). No significant within-subject effect of Test Rule Type was observed, $F(1, 30) = 1.324$, $p = 0.259$, *partial eta squared* = 0.042. The between-subject effect of Type Frequency was not significant, $F(1, 30) = 0.645$, $p = 0.428$, *partial eta squared* = 0.021. The interaction between Test Rule Type and Type Frequency reached significance, $F(1, 30) = 4.271$, $p = 0.048$, *partial eta squared* = 0.125. These results suggest that infants in Experiment 5 and in Experiment 6 responded differently to the trained and the non-trained rules.

As suggested by these supplementary ANOVA analyses, the effect of type frequency on learning of word order rules was more clearly pronounced in experiments without morphological markings. In the ANOVA conducted on the results of these two experiments (Experiments 5 and 6), the interaction between Test

Rule Type and Type Frequency was significant. Whereas in the case of identical experiments with morphological markings (Experiments 3 and 4), that interaction did not reach significance. These results suggest that infants learned the rule from the training with a high type frequency of rule examples better when stimuli were morphologically unmarked.

To further examine whether learning in unmarked experiments (Experiments 5 and 6) was indeed better than in morphologically marked experiments (Experiments 3 and 4), we performed a Mixed-design ANOVA on the results of Experiments 3, 4, 5 and 6. The dependent variable was infants' looking time during test trials. The within-factor was Test Rule Type (trained vs. non-trained). There were two between-factors: Type Frequency of Rule Instances (80% in Experiments 3 and 5, and 50% in Experiments 4 and 6) and Morphological Markings (marked in Experiments 3 and 4, and unmarked in Experiments 5 and 6). The within-subject effect of Test Rule Type was nearly significant, $F(1, 60) = 3.827$, $p = 0.055$, *partial eta squared* = 0.06. The between-subject effect of Type Frequency was not significant, $F(1, 60) = 0.366$, $p = 0.547$, *partial eta squared* = 0.006. The between-subject effect of Morphological Markings was not significant either, $F(1, 60) = 1.381$, $p = 0.245$, *partial eta squared* = 0.023. The interaction between Test Rule Type and Morphological Markings was not significant, $F(1, 60) = 0.434$, $p = 0.513$, *partial eta squared* = 0.007. The interaction between Test Rule Type, Type Frequency and Morphological Markings was not significant either, $F(1, 60) = 0.138$, $p = 0.712$, *partial eta squared* = 0.002. These results did not support the alternative hypothesis that less morphology enhances learners' attention to the word order and leads to the better learning of the rule. Note that our initial hypothesis was the contrary, that morphological markings may support learning. According to the ANOVA results, morphological markings did not impede nor support the learning of word order movement rules.

Consistent with our main hypothesis about the role of type frequency for rule learning, the ANOVA described above (for Experiments 3, 4, 5 and 6) revealed a significant interaction between Test Rule Type and Type Frequency, $F(1, 60) = 7.299$, $p = 0.009$, *partial eta squared* = 0.108. To further examine this interaction, we conducted two separate analyses. First, we performed a Paired-samples *t*-test on infants' looking times in two experiments where the type frequency of rule examples was dominant (i.e., the results of Experiments 3 and 5 combined together). This Paired-samples *t*-test showed a significant difference of infants' looking times for the trained and the non-trained rules, $t(31) = -3.74$, $p = .001$, *two-tailed, partial eta squared* = 0.311. Infants were looking longer for the non-trained rule ($M = 9.17$, $SE = 0.83$) than for the trained one ($M = 7.07$, $SE = 0.72$). This suggests that in Experiments 3 and 5 where the type frequency of rule examples in the training was dominant relative to noise examples, infants learned the rule and applied it to novel instances. In the second analysis, we conducted a Paired-samples *t*-test on infants' looking times in two experiments where rule examples were not dominant by type in the training (i.e., the results of Experiments 4 and 6 combined together). The analysis showed no significant difference of looking times for the trained ($M = 7.69$, $SE = 0.84$) and the non-trained ($M = 7.35$, $SE = 0.68$) rules, $t(31) = 0.49$, $p = 0.631$, *two-tailed, partial eta squared* = 0.008. This suggests that infants in Experiments 4 and 6 did not discriminate between the trained and the non-trained rules. Hence, the interaction between Test Rule Type and Type Frequency observed in the ANOVA conducted on the results of Experiments 3, 4, 5 and 6 was driven by the difference of looking times for the trained and the non-trained rules in Experiments 3 and 5, where the type frequency of rule examples was dominant in the input.

An important question in language acquisition is what are the contributions of innate universal grammar principles (Chomsky, 1986) and distributional learning. This thesis did not address the contribution of universal grammar principles. Abstract rules used here were artificial rules designed with two word order movement patterns.

These patterns were available to learners on a surface and did not require learning of deeper syntactic structures. In this work, the focus was on the contribution of distributional properties of the input to learning. The constraints shown by type and token frequencies might be an innate mechanism of analyzing surface forms of utterances. But this is not the crucial test for Chomskyan UG.

Recent work by Lidz and colleagues (Lidz, Gleitman and Gleitman, 2003; Lidz and Gleitman, 2004) has shown that surface characteristics of language input do not have equal impact on children's mapping between semantics and syntax. Lidz, Gleitman and Gleitman (2003) used causativity in the Kannada language as a testing case. In Kannada, causativity has two characteristics in the input: a causative morpheme and the transitivity requirement. The verb affix '*isu*' is the causative marker. It is obligatory for most of causative verbs (e.g., *MoSale kudure-yannu eer-is-utt-ade – Alligator horse-ACC rise-CAUS-NPST-3SN – The alligator raises the horse*). However, there is a small set of causative verbs in Kannada that do not require the causative morpheme. Besides, causativity in Kannada is always expressed with transitive structures. Transitivity requires a direct object, so that sentences with transitive verbs typically contain at least two NPs (that correspond to two arguments). In the example described above (*MoSale kudure-yannu eer-is-utt-ade – Alligator horse-ACC rise-CAUS-NPST-3SN – The alligator raises the horse*) the transitivized causative verb '*raise*' is used with two NPs: '*the alligator*' and '*the horse*'. Intransitive structures with one NP cannot be causative. On the other hand, transitive structures are not always causative, since many verbs are transitive but not causative. For example, the transitive sentence *MoSale kudure-yannu nooD-utt-ade – Alligator horse-ACC see-NPST-3SN – The alligator sees the horse* has the non-causative verb and two NPs. According to the authors, the number of NPs is only a probabilistic cue to causativity since many transitive verbs are not causative. In contrast, the authors considered the causative morpheme as a much more reliable cue to causativity. Therefore they tested the weighting of these two cues in children's interpretation of causativity, to determine whether children will rely on the stronger morphological cue.

In their study, three-year-old Kannada-speaking children were asked to act-out utterances with toy animals. There were four kinds of utterances: NP V, NP Vcaus, NP NPacc V and NP NPacc Vcaus. NP V sentences contained intransitive verbs (transitive verbs are not grammatical in this structure in Kannada, similarly to English): e.g., *Simha hoog-utt-ade – Lion go-NPST-3SN – The lion goes.* NP Vcaus sentences contained transitive verbs. This is ungrammatical in Kannada. The ungrammaticality of NP Vcaus utterances was of two kinds. In some sentences, the second NP required by the transitive structure was absent (e.g., **Jinke naDug-is-utt-ade – Deer shake-CAUS-NPST-3SN – The deer shakes*). In other sentences the causative morpheme was added to non-causative verbs (e.g., **Huli hoog-is-utt-ade – Tiger go-CAUS-NPST-3SN – The tiger goes*). As for the NP NPacc V structure the authors made grammatical sentences such as *Miinu mosale-yannu muTT-utt-ade – Fish alligator-ACC touch-NPST-3SN – The fish touches the alligator*, and ungrammatical sentences such as **Kothi yeth-annu eer-utt-ade – Monkey ox-ACC rise-NPST-3SN – The monkey raises the ox*. The latter example was made with the intransitive verb *eeru* (‘*rise*’) which can be transitivized by adding the causative morpheme. In this example, the causative morpheme was ungrammatically absent and hence the verb *eeru* stayed intransitive. The last structure used in the stimuli - NP NPacc Vcaus – normally requires the use of intransitive verbs (like *eeru* (‘*rise*’) mentioned previously. Once these verbs receive the causative marking, they become transitive and require the second NP. For example, one of grammatical sentences used for stimuli in this structure was: *Huli jinke-yannu kuN-is-utt-ade – Tiger deer-ACC jump-CAUS-NPST-3SN – The tiger makes the deer jump*). In this example, an intransitive verb ‘*jump*’ was used in a transitive structure due to the added causative morpheme. Other sentences in the NP NPacc Vcaus structure were ungrammatical (e.g., **Yethu ghenda mruga-vannu jigut-is-utt-ade – Ox rhinoceros-ACC pinch-*

CAUS-NPST-3SN – The ox makes the rhinoceros pinch). In this case, the transitive verb '*pinch*' receives a causative marking, which is grammatical in Kannada but has the effect of adding the triadic meaning requiring three arguments. In the above example, the third argument was ungrammatically absent.

The utterances in these four kinds of structures differed by the number of noun phrases: one or two. At a deeper level, this difference corresponded to a different number of arguments (one or two). The utterances used in the stimuli also differed by whether the verb was marked with a causative morpheme. Across the one-NP and two-NP structures, the verbs were either marked or not. This design allowed testing which cue would be used by children to infer causativity.

Children's actions in the act-out task were labeled as causative and non-causative. Their performance was analyzed in the ANOVA with three factors: the number of arguments (one or two; this number was identical to the number of noun phrases in the sentence), morphology (bare or causative) and valency (intransitive or transitive). The results suggested that three-year-olds interpreted verbs as causative relying strongly on the number of arguments (or noun phrases) in the sentence and ignoring the morphological form of verbs. According to Lidz, Gleitman and Gleitman (2003), this preference goes against the fact that the morphological cue is the more reliable cue of causativity in Kannada. The authors considered the number of NPs as the number of arguments not available at the surface of utterances, unlike the morphological cue. Goldberg (2004) argued that both the morphological cue and the cue of the number of NPs are available at the surface. Regardless of these differential views, the less distributionally reliable cue of the number of NPs outweighed the other, more reliable, morphological cue. Children's tendency to rely on the number of NPs as a cue to causativity was also previously observed in English. In English, causative meaning does not require any particular morpheme. But the relation of causativity to transitivity, and hence, the number of NPs is similar to the Kannada language. In English, transitivity is also a probabilistic cue to causativity. Causative verbs cannot be intransitive. For example, an intransitive verb *run* does not allow causativity: *John runs*. Causative verbs can only be transitive (e.g., *John raises the curtains*). On the other hand, transitive verbs are not necessarily causative. For example, a transitive verb *like* does not have a causative meaning: *John likes the toys*. Lidz, Gleitman and Gleitman (2003) cite a previous study by Fisher (1996) where English-speaking children interpreted sentences with nonsense verbs as causative when the sentence contained a transitive structure, that is, when it contained a direct object as the second NP. Similarly, they interpreted nonsense verbs as non-causative when the structure of the sentence was intransitive, that is, if the sentence had only one NP. These results suggested that children had a bias for interpreting two NP sentences as causative. In English, such bias is supported by distributional characteristics of the input, since there exists no better cue to causativity than the probabilistic cue of the number of NPs. In Kannada, however, such distributionally reliable cue exists: the causative morpheme. However, Kannada-learning children ignored the more reliable morphological cue in their interpretation of causativity. Lidz, Gleitman and Gleitman (2003) interpreted these results as the evidence for an innate bias in children to rely on the number of NPs to infer the causative meaning.

Our work did not address the question of mapping between semantics and syntax. It only revealed effects of distributional properties of learning input available on a surface. Our study did not test the innate knowledge in infants.

Our results suggest that infants track both distribution of general patterns in the input and learn specific instances used in those patterns. This interpretation can be derived from results of Experiments 3 and 13 where the training was identical. Infants learned the trained rule and generalized it to novel instances (in Experiment 3) and also resisted applying that rule to noise sentences (in Experiment 13). The last results suggest that infants tracked those noise sentences from the training. Such capacity of learners to track both general and specific statistics in the input was previously shown in studies with adults and children (Wonnacott, Newport & Tenenbaum, 2008; Wonnacott, 2011). The study of Wonnacott, Newport & Tenenbaum (2008) was modeled in a probability-based framework of Bayesian learning. Perfors, Tenenbaum & Wonnacott (2010) used a Hierarchical Bayesian model to simulate the learning from such input. They obtained the same results as Wonnacott, Newport & Tenenbaum (2008) where type frequency supported learning of general patterns. However, Wonnacott, Newport & Tenenbaum (2008) did not separate type frequency from overall frequency, so it is not clear whether learners relied more on one of those frequencies. It is therefore not clear whether the Hierarchical Bayesian model predicts any difference in the role of type and overall frequencies. Although Wonnacott, Newport and Tenenbaum (2008) showed the effect of token frequency on learning of specific instances, that experimental condition was not modeled by Perfors, Tenenbaum and Wonnacott (2010). Perfors, Tenenbaum and Wonnacott (2010) applied the Hierarchical Bayesian model to a different experimental condition of Wonnacott, Newport and Tenenbaum (2008), where adults showed learning of specific instances from only four repetitions. In that condition, it was not shown whether learning of specific instances could fail. Hence, it is not clear whether the Hierarchical Bayesian model predicts that low token frequency would impede learning of utterances as specific instances (as in the learning of exceptions).

Our findings show not only the distributional conditions favorable for generalization and learning of exceptions, but also the distributional conditions under which these processes would be impeded. The Hierarchical Bayesian model developed by Perfors, Tenenbaum and Wonnacott (2010) did not specifically address this issue.

Besides, the learning process indicated by our results is generally different from Bayesian learning assumptions. Our results suggest that infants learned or failed to learn the rule based on the properties of the input. In particular, a high number of sentence types not conforming to the rule impeded learning. The question is whether infants learned those non-rule-conforming sentences (i.e., the noise) as another pattern. In our experiments, the non-rule-conforming examples were sentences with an ABC order that did not go through the word order movement rule (note that the rule was ABC-BAC for one group of infants and ABC-ACB for another group of infants). It is possible that the noise itself was the pattern, and that both rule and noise patterns were learned. If indeed infants learned both rule and noise patterns when they were equally available in the input, we should have observed learning in Experiments 4 and 6. In these two experiments both rule and noise patterns were equally available in the training in 50% of cases. After the training, infants were tested with the trained and untrained movement rules. If they had learned all the patterns available in the input, they should have discriminated the trained rule from the untrained rule. However, we did not find evidence of such discrimination. It can be concluded that they did not learn the movement pattern in the training. The results of Experiments 4 and 6 suggest that infants treated the non-application cases as noise. However, according to Bayesian learning, both rules should be learned if they were equally available in the input. Moreover, learners should make their choice of the most probable rule by comparing the input with all possible rules that could be derived from its components. For example, upon hearing three syllable sequences ABA (like in Gerken, 2006), learners should consider the probability of such rules as AAA, BBB, ABB etc., even though they are not present in the input. If the input consistently supports the ABA rule, learners choose it as the most probable one, and all other rules are rejected. Such assumption has been often criticized for psychological implausibility (e.g., Endress, 2013). Criticizing one of Bayesian learning models, Endress (2013) wrote: "Taking Frank and Tenenbaum's (2011) model at face value, they claim that, once infants enter an experimental room, they keep track of all syllables they have heard in the experiment, and while comfortably seated on their parent's lap, contemplate all possible sequences that can be formed with these syllables, as well as all possible rules with which each of these hypothetical sequences might or might not be consistent" (pp. 161-162). Bayesian learning models are 'ideal observer' models. These models assume that learners acquire all possible rules from the input and then choose among them. This assumption does not take into consideration constraints in children's processing.

In our study, infants did not treat the noise cases as another pattern. This might be related to the fact that the noise was made with non-application sentences. Future research can examine whether learning will be the same from the input where the noise sentences are overt violations of the rule. For example, infants can be trained with sentences conforming to two rules: ABC-BAC and ABC-ACB. After such training, one group of infants can be tested with novel sentences conforming to one of the trained rules, ABC-BAC, and an untrained rule, ABC-CBA. Another group of infants can be tested with the second trained rule, ABC-BAC, and the untrained rule, ABC-CBA. If infants' learning is similar to Bayesian learning, they should learn both rules that are equally available in the training.

Although our results are fundamentally different from the assumptions of Bayesian learning, some of our findings are compatible with some predictions of Bayesian models. In particular, the principle of Bayesian learning suggests that learners choose the more restrictive pattern when a learning input gives equal evidence for multiple patterns (Frank & Tenenbaum, 2011). The preference for more restrictive patterns was shown in Experiments 3 and 11 where the training stimuli allowed two possible generalizations. The more restrictive generalization required obligatory morphological markings to apply the word order movement rule, while the broader generalization allowed applying the word order movement rule to novel instances irrespective of their morphological markings. In Experiment 11 we did not find evidence of infants' broader generalization. On the contrary, the results of

Experiments 3 and 11 suggest that the generalization made by infants was the restrictive one.

In conclusion, this thesis has demonstrated the important role of type and token frequencies in infants' generalization of rules and their learning of exceptions. When rule instances were high in type frequency relative to noise, infants learned the rule and generalized it to novel instances. When rule and noise instances in the training were equal in type frequency, generalization was impeded. We also found that infants make a more conservative generalization when the input allows two generalizations, one broader and more abstract, another more narrow and conservative. It was also found that over-generalization and the learning of exceptions can be distributionally based. When noise exemplars in the input had low token frequency, infants generalized them to the more frequent rule. These results showed over-generalization. On the other hand, when noise exemplars in the input had high token frequency, infants resisted over-generalizing them to the rule, suggesting the learning of exceptions.

APPENDIX A

SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 1 AND 2

1. Dozhd' zalil cherdak – Rain flood attic – Rain has flooded the attic
2. Veter gnjot derev'ja – Wind bend trees – The wind is bending trees
3. Ozero pokrylos' l'dom – Lake get covered ice – The lake got covered with ice
4. Koni otdokhnut segodnja – Horses rest today – The horses will rest today
5. Staja letit klinom – Flock fly wedge – The flock is flying in a wedge
6. Vorona nashla pugovitsy – Crow find buttons – The crow has found buttons
7. Belka zapaslas' orehami – Squirrel stock up nuts – The squirrel has stocked up on nuts
8. Solntse ozarilo gostinuju – Sun lit up living-room – The sun lit up the living-room

All sentences went through the ABC -> BAC movement for training one group of infants, and through the ABC -> ACB movement for training another group of infants.

APPENDIX B

SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENTS 1 AND 2

1. Sneg ukutal gorod – Snow muff up city – Snow has muffled up the city
2. Druz'ja prinesli shhenka – Friends bring puppy – Friends brought a puppy
3. Inej morozit stjokla – Frost freeze windows – Frost is freezing over the windows
4. Medved' naelsja mjoda – Bear make a feast honey – The bear has made a feast of the honey

Sentences 1 and 2 went through the ABC –> BAC movement; sentences 3 and 4 went through the ABC –> ACB movement.

APPENDIX C

SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENTS 3, 4 AND 7

1. Njura topit pechku – Njura stoke furnace – Njura is stoking a furnace
2. Tjoma rubit lipku – Tjoma hack linden – Tjoma is hacking a linden

Each sentence went through the ABC –> BAC and ABC –> ACB movements, depending on the type of trial and the group of infants.

The marking *–a* is a noun inflection for nominative case singular; the marking *–it* is a verb inflection for the third person singular; the marking *–ku* consists of a suffix *–k* and an inflection *–u*; *–u* is a noun inflection for accusative case singular. The suffix *-k* is a diminutive suffix, although for some words the *-k* suffix no longer has the diminutive meaning, probably because the form fossilized during historical evolution and lost the diminutive meaning.

APPENDIX D

RULE-CONFORMING SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 3, 4, 7, 9, 11, 13 AND 14

1. Vika darit murku – Vika give kitty – Vika is giving a kitty (as a present)
2. Dima gonit galku – Dima chase away jackdaw – Dima is chasing away a jackdaw
3. Lera nosit dochku – Lera carry daughter – Lera is carrying her daughter
4. Njuta varit kashku – Njuta cook porridge – Njuta is cooking porridge
5. Gosha lovit rybku – Gosha catch fish – Gosha is catching fish
6. Vova katit lodku – Vova roll boat – Vova is rolling a boat
7. Rada manit koshku – Rada lure cat – Rada is luring a cat
8. Zhora lepit belku – Zhora sculpt squirrel – Zhora is sculpting a squirrel

All sentences went through the ABC -> BAC movement for training one group of infants, and through the ABC -> ACB movement for training another group of infants.

APPENDIX E

NOISE SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 3, 11, 12, 13 AND 14

1. Gena vidit lavku – Gena see bench – Gena sees a bench
2. Lida zharit manku – Lida fry semolina – Lida is frying semolina

These sentences did not go through any movement.

APPENDIX F

NOISE SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 4, 7, 9 AND 10

1. Gena vidit lavku – Gena see bench – Gena sees a bench
2. Lida zharit manku – Lida fry semolina – Lida is frying semolina
3. Kesha penit rechku – Kesha foam river – Kesha is foaming the river
4. Ljova kopit lesku – Ljova store up fishing line – Ljova is storing up some fishing line
5. Nina lechit sivku – Nina treat horse – Nina is treating a horse
6. Goga vozit zhuchku – Goga transport doggy – Goga is transporting a doggy
7. Sasha parit repku – Sasha steam turnip – Sasha is steaming a turnip
8. Seva kosit gorku – Seva mow hill – Seva is mowing a hill

These sentences did not go through any movement.

APPENDIX G

SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENTS 5, 6, 8, 9, 10, 11 AND 12

1. Snova milyj vesel – Again darling happy – Again the darling is happy
2. Vizhu nosik belki – See little nose squirrel – (I) can see a little nose of a squirrel

Each sentence went through the ABC –> BAC and ABC –> ACB movements, depending on the type of trial and group of infants.

APPENDIX H

RULE-CONFORMING SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 5, 6, 8 AND 15

1. Machty gnutsja lukom – Masts bend bow – The masts are bending in a bow
2. Zina gladit plat'e – Zina iron dress – Zina is ironing a dress
3. Pojte pesnju druzhno – Sing song together – Sing a song together
4. Veter vybil okna – Wind knock out windows – The wind has knocked out the windows
5. Dimke snilos' pole – Dimka dream field – Dimka was dreaming of a field
6. Chistim tufli vaksoj – Clean shoes polish – (We) are cleaning the shoes with polish
7. Budesh vilkoj kushat' – Will fork eat – (You) will be eating with a fork
8. Flagi utrom snjali – Flags morning take down – The flags were taken down in the morning

All sentences went through the ABC –> BAC movement for training one group of infants, and through the ABC –> ACB movement for training another group of infants.

APPENDIX I

NOISE SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 5 AND 15

1. Stanut reki polny – Become rivers full – The rivers will become full
2. Otzvuk smekha sladok – Echo laugh sweet – The echo of a laugh is sweet

These sentences did not go through any movement.

APPENDIX J

NOISE SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 6 AND 8

1. Stanut reki polny – Become rivers full – The rivers will become full
2. Otzvuk smekha sladok – Echo laugh sweet – The echo of a laugh is sweet
3. Kozam travok ssyplju – Goats grass sack – (I) will sack some grass to goats
4. Seno pahnet volej – Hay smell freedom – Hay smells with freedom
5. Skrojut tuchi solntse – Hide clouds sun – The clouds will hide the sun
6. Tanets veren bubnu – Dance be true tambourine – The dance is true to tambourine
7. Pishem bukvy krasnym – Write letters red – (We) are writing the letters with the red
8. Obuv' skinul rezvo – Shows throw off quickly - (He) has thrown off the shoes quickly

These sentences did not go through any movement.

APPENDIX K

RULE-CONFORMING SENTENCES WITH AN ABC ORDER USED IN THE TRAINING FOR EXPERIMENTS 10 AND 12

1. Vika darit murku – Vika give kitty – Vika is giving a kitty (as a present)
2. Dima gonit galku – Dima chase away jackdaw – Dima is chasing away a jackdaw
3. Gosha lovit rybku – Gosha catch fish – Gosha is catching fish
4. Vova katit lodku – Vova roll boat – Vova is rolling a boat
5. Rada manit koshku – Rada lure cate – Rada is luring a cat
6. Zhora lepit belku – Zhora sculpt squirrel – Zhora is sculpting a squirrel
7. Pojte pesnju druzhno – Sing song together – Sing the song together
8. Veter vybil okna – Wind knock out windows – The wind has knocked out the windows

All sentences went through the ABC -> BAC movement for training one group of infants, and through the ABC -> ACB movement for training another group of infants.

APPENDIX L

SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENTS 13 AND 14

1. Gena vidit lavku – Gena see bench – Gena sees a bench
2. Lida zharit manku – Lida fry semolina – Lida is frying semolina

Each sentence went through the ABC –> BAC and ABC –> ACB movements, depending on the type of trial and the group of infants.

APPENDIX M

SENTENCES WITH AN ABC ORDER USED IN THE TEST FOR EXPERIMENT 15

1. Stanut reki polny – Become rivers full – The rivers will become full
2. Otzvuk smekha sladok – Echo laugh sweet – The echo of a laugh is sweet

Each sentence went through the ABC –> BAC and ABC –> ACB movements, depending on the type of trial and the group of infants.

APPENDIX N

INFANTS' EXPOSURE TO LANGUAGES ACCORDING TO PARENTAL REPORT

The number of infants with an exposure to a particular language is given in brackets.

Experiment 1:

> 70% French (14), > 70% English (1), 60% Spanish 30% French 10 % English (1)

Experiment 2:

> 70% French (11), > 70% English (1), 50% French 50% German (1), 50% French 50% Mandingo (1), 50% French 50% Portuguese (1), 50% Japanese 30% French 20% English (1)

Experiment 3:

> 70% French (10), > 70% English (2), > 70% Mandarin (1), 60% French 40% Arab (1), 50% English 50% French (1), 50% French 50% Spanish (1)

Experiment 4:

> 70% French (9), > 70% English (1), > 70% Portuguese (1), > 70% Spanish (1), 50% English 50% French (2), 50% French 50% Spanish (2)

Experiment 5:

> 70% French (9), > 70% Mandarin (1), > 70% Portuguese (1), 65% French 30% German 5% Arab (1), 60% Berber 40% French (1), 50% English 50% French (1), 50% French 16.6% English 16.6% Portuguese 16.6% Spanish (1), 45% English 45% French 10% Spanish (1)

Experiment 6:

> 70% French (14), > 70% English (1), > 70% Spanish (1)

Experiment 7:

> 70% French (13), > 70% English (2), 50% English 50% French (1)

Experiment 8:

> 70% French (12), 60% English 40% French (1), 60% French 20% English 20% Haitian Creole (1), 50% English 30% French 20% Portuguese (1), 50% French 50% Spanish (1)

Experiment 9:

> 70% French (12), > 70% Arab (1), 60% English 40% French (2), 50% French 45% English 5% Portuguese (1)

Experiment 10:

> 70% French (9), > 70% English (2), 60% English 40% French (1), 50% French 50% Spanish (2), 50% Arab 50% French (1), 33.3% Corean 33.3% English 33.3% French (1)

Experiment 11:

> 70% French (11), 65% Spanish 35% French (1), 60% Arab 40% Italian (1), 60% Spanish 30% English 10% French (1), 50% French 40% English 8% Greek 2% Haitian Creole (1), 45% French 35% Cantonese 20% Spanish (1)

Experiment 12:

> 70% French (13), > 70% Japanese (1), 50% German 40% Spanish 10% English (1), 45% English 45% Spanish 10% Cantonese (1)

Experiment 13:

> 70% French (10), > 70% English (3), > 70% Arab (1), > 70% Spanish (1), 60% Italian 40% Slovac (1)

Experiment 14:

> 70% French (11), > 70% English (2), 65% French 35% English (1), 55% French 45% Arab (1), 50% English 50% French (1)

Experiment 15:

> 70% French (12), 60% French 30% Spanish 10% English (1), 60% French 25% German 15% English (1), 50% French 50% Laosian (1), 50% French 50% Spanish (1)

REFERENCES

Ambridge, B., Pine, J.M., Rowland, C.F. and Young, C.R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition, 106*(1), 87-129.

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry, 10*(4), 533-581.

Bowerman, M. (1974). Learning the structure of causative verbs: A study in the relationship of cognitive, semantic, and syntactic development. *Papers and Reports on Child Language Development, 8*, 142-178.

Bowerman, M. (1982a). Evaluating competing linguistic models with language acquisition data: Implications of developmental errors with causative verbs. *Quaderni di Semantica, 3*, 5-66.

Bowerman, M. (1982b). Reorganizational processes in lexical and syntactic development. In E. Wanner and L.R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 319-346). New York: Academic Press.

Bowerman, M. (1983). How do children avoid constructing an overly general grammar in the absence of feedback about what is not a sentence? *Papers and Reports on Child Language Development, 22*, 23-35.

Bowerman, M. (1988). The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In J. Hawkins (Ed.), *Explaining language universals* (pp. 73-101). Oxford: Basil Blackwell.

Braine, M. D. S. (1971). On two types of models of the internalization of grammars. In D. I. Slobin (Ed.), *The ontogenesis of grammar* (pp. 153-186). New York: Academic Press.

Braine, M. D. S., and Brooks, P. J. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello and W. Merriman (Eds.), *Beyond names of things: Young children's acquisition of verbs* (pp. 352–376). Hillsdale, NJ: Erlbaum.

Brooks, P. J., and Tomasello, M. (1999). How children constrain their argument structure constructions. *Language, 75*, 720–781.

Brooks, P. J., Tomasello, M., Dodson, K. and Lewis, L. (1999). Young children's overgeneralizations with fixed transitivity verbs. *Child Development, 70*, 1325–1337.

Brooks, P. J., and Zizak, O. (2002). Does pre-emption help children learn verb transitivity? *Journal of Child Language, 29*, 759–781.

Chomsky, N. (1986). *Knowledge of language: Its structure, origins and use.* New York: Praeger.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1-33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, L. B., Atkinson, D. J. and Chaput, H. H. (2000). Habit 2000: A new program for testing infant perception and cognition (Version 2.2.5c) [Computer software]. Austin: University of Texas.

Cyr, M., and Shi, R. (2013). Development of abstract grammatical categorization in infants. *Child Development, 84*(2), 617-629.

Endress, A. D. (2013). Bayesian learning and the psychology of rule induction. *Cognition, 127*, 159-176.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology, 31*(1), 41 – 81.

Frank, M. C., & Tenenbaum, J. B. (2011). Three ideal observer models for rule learning in simple languages. *Cognition, 120*, 360-371.

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

Gerken, L. A. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition, 115*, 362–366.

Gerken, L. A., Wilson, R. and Lewis, W. (2005). 17-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249–268.

Göksun, T., Küntay, A. and Naigles, L. (2008). Turkish children use morphosyntax in extending verb meaning. *Journal of Child Language, 35*, 291-323.

Goldberg, A. E. (1995). *A construction grammar approach to argument structure.* Chicago: The University of Chicago Press.

Goldberg, A. E. (2004). But do we need Universal Grammar? Comment on Lidz, Gleitman and Gleitman (2003). *Cognition, 94*, 77–84.

Goldsmith, J. (1980). Meaning and mechanism in grammar. In S. Kuno (Ed.), *Harvard studies in syntax and semantics, 3* (pp. 423–449). Cambridge, MA: Harvard University Press.

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

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

Gómez, R. L., and LaKusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7*, 567–580.

Green, G. M. (1974). *Semantics and syntactic regularity.* Bloomington, Indiana University Press.

Hochberg, J. G. (1986). Children's judgements of transitivity errors. *Journal of Child Language, 13*, 317-334 .

Hudson, C.L., and Newport, E.L. (1999). Creolization: Could adults really have done it all? In A. Greenhill, H. Littlefield, and C. Tano (Eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development: Vol. 1* (pp. 265-276). Somerville, MA: Cascadilla Press.

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

Höhle, B., Schmitz, M., Santelmann, L. and 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.

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

Johnson, S. P., Fernandas, K. J., Frank, M. C., Kirkham, N. Z., Marcus, G. F., Rabagliati, H. and Slemmer, J. A. (2009). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy, 14*, 2-18.

Lidz, J., Gleitman, H., and Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition, 87*, 151 – 178.

Lidz, J., and Gleitman, L. (2004). Yes, we still need universal grammar: reply. *Cognition, 94*, 85–93.

Maratsos, M., Gudeman , R., Gerard-Ngo, P. and DeHart , G. (1987). A study in novel word learning: The productivity of the causative. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 221-256). Hillsdale, NJ: Erlbaum.

Marcus, G. F., Fernandes, K. J. and Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18*, 387-391.

Marcus, G. F., Vijayan, S., Bandi Rao, S. and Vishton, P. M. (1999). Rule-learning in seven-month-old infants. *Science, 283*, 77-80.

Markman, E. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

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

Mazurkewich, I., and White, L. (1984). The acquisition of the dative alternation: Unlearning overgeneralizations. *Cognition, 16*, 261-283.

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

Mintz, T. H. (2006). Finding the verbs: Distributional cues to categories available to young learners. In K. Hirsh-Pasek and R. M. Golinkoff (Eds.), *Action Meets Word: How Children Learn Verbs* (pp. 31-63). New York: Oxford University Press.

Morgan, J. L., and Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior, 20*, 67–85.

Newport, E. L. (1999). Reduced input in the acquisition of signed languages: Contributions to the study of creolization. In M.Degraff (Ed.), *Creolization, diachrony, and language acquisition* (pp. 161–78). Cambridge, MA: MIT Press.

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

Oehlre, R. (1976). *The grammatical status of the English dative alternation*. Doctoral dissertation, Cambridge, MA: MIT.

Oshima-Takane, Y., Ariyama, J., Kobayashi, T., Katerelos, M. and Poulin-Dubois, D. (2011). Early verb learning in 20-month-old Japanese-speaking children. *Journal of Child Language, 38*, 455-484.

Peña, M., Bonatti, L. L., Nespor, M. and Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604-607.

Perfors, A., Tenenbaum, J.B. and Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language, 37*, 607-642.

Pinker, S. (1984). *Language learnability and language development.* Cambridge: Harvard University Press.

Pinker, S. (1987). *Resolving a learnability paradox in the acquisition of the verb lexicon.* Cambridge, MA: Lexicon Project.

Pinker, S., Lebeaux, D. S. and Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition, 26*, 195-267.

Ross, D. S., and Newport, E. L. (1996). The development of language from non-native linguistic input. In A. Stringfellow, D. Cahana-Amitay, E. Hughs, and A. Zukowski (Eds.), *Proceedings of the 20th annual Boston University Conference on Language Development* (pp. 623–645). Boston: Cascadilla.

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

Saffran, J.R., Pollak, S.D., Seibel, R.L. and Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition, 105*, 669-680.

Santelmann, L., and Jusczyk, P. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for processing limitations. *Cognition, 69*, 105-134.

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

Singleton, J.L., and Newport, E.L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology, 49*, 370-407.

Stowell, T. (1981). *Origins of phrase structure*. Doctoral dissertation. Cambridge, MA: MIT.

Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance limitations on grammaticality judgment tasks. *Cognitive Development, 19*, 15–34.

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

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

Waxman, S. R., and Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology, 43*(3), 217-242.

Waxman, S.R., Lidz, J., Braun, I. E. and Lavin, T. (2009). Twenty-four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology, 59*(1), 67-95.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language, 65*, 1-14.

Wonnacott, E., and Newport, E.L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In A. Brugos, M.R. Clark-Cotton, and S. Ha (Eds.), *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 663-673). Somerville, MA: Cascadilla Press.

Wonnacott, E., Newport, E.L. and Tanenhaus, M.K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*, 165-209.